Analysing soil organic carbon using near infrared spectroscopy in the laboratory and in the field on-line

Von der
Fakultät Architektur, Bauingenieurwesen und Umweltwissenschaften
der Technischen Universität Carolo-Wilhelmina
zu Braunschweig

zur Erlangung des Grades einer
Doktorin der Naturwissenschaften (Dr. rer. nat.)
genehmigte

Dissertation

von
Jasmin Aymone Miltz
geboren am 18.11.1975
aus Celle

Eingereicht am:	21. Januar 2016
Disputation am:	09. Mai 2016
Berichterstatter:	Prof. Dr. R. Nieder, Technische Universität Braunschweig
	Prof. Dr. H. Flessa, Georg August Universität Göttingen

2016

Table of contents

I List of tables

II List of figures

III List of abbreviations

1^{st} der	First Savitzky-Golay derivative
2^{nd} der	Second Savitzky-Golay derivative
AIC	Akaike value
ANOVA	Analysis of variance
ASD	Analytical Spectral Devices
BD	Bulk density
BD_{CALL}	BD calculated out of PTFs mentioned in Callesen et al. (2003)
BD_{FS}	BD calculated out of field-specific data: SOC data from direct chemical analysis and BD data (BD_{MEAS} where available, otherwise BD_{CALL}).
BD_{MEAS}	BD calculated out of reference SOC data measured via direct chemical analysis
CALL	PTF as published in Callesen et al. (2003)
C	Carbon
CC	Carbon – Carbon bond
CH	Methine group
CN	Cyanide group
CO	Carbonyl group
CO_2	Carbon dioxide
COV	Covariance
CV	Coefficient of variation
DIN	Deutsches Institut für Normung
EC	Electrical conductivity
FAO	Food and Agriculture Organization
FIELDSPEC	Fieldspecific
FIR	Far infrared
FT	Fourier Transform
GPS	Global Positioning System
h	hour
ha	hectar
IQ	Interquartile range
MC	Monte Carlo simulation

MDD	Minimum detectable difference
MIR	Mid infrared
MPA	Multi purpose analyser
n	number
N	Nitrogen
NH	Amino group
NIR	Near infrared
NIRS	Near infrared spectroscopy
NM	Not mentioned
Normal	Normalisation
OBK	Organischer Bodenkohlenstoff
OH	Hydroxyl group
OM	Organic matter
p	Density
PbS	Lead-sulfide
PLS	Partial least squares
PLSR	Partial least squares regression
PRED	Predicted
PTF	Pedotransfer function
R	Correlation coefficient
R^2	Coefficient of determination
RA	Research assistant
Refl	Reflectance
Rel.	Relative
RER	Range of error ratio
RMSECV	Root mean square error of cross validation
RMSEP	Root mean square error of prediction
RPD	Ratio of prediction to deviation
RPIQ	Ratio of performance of IQ
s	second
SD	Standard deviation
SE	Standard error
SE_{MC}	SE of Monte Carlo simulation

SE_{MEAS}	SE/ measurement error (calculated out of error of BD and SOC data and covariance between SOC and BD data)
SE_{PRED}	SE of predicted SOC values (received via off-line prediction)
SE_{REF}	SE of reference SOC values (measured via direct chemical analysis)
SE_{RMSEP}	SE calculated out of calibration error (RMSECV ~ RMSEP)
SEC	Standard error of calibration
SIC	Soil inorganic carbon
SNV	Standard normal variate
SOC	Soil organic carbon
STA	Student assistant
TA	Technical assistant
UNFCCC	United Nations Framework Convention on Climate Change
VIS	Visible part of the electromagnetic spectrum
y_i	measured reference values
$\hat{y}_i$	predicted values out of cross-validation
yr	year
σBD	SE of BD
σSOC	SE of SOC

IV Abstract

Soil organic carbon (SOC), as a key property of soil quality maintenance, varies over space and time. The assessment and monitoring of SOC is important to ensure sustainable soil management. SOC can be determined by conventional laboratory analytical techniques, but the preparation and measurement of numerous soil samples can be costly. Near infrared spectroscopy (NIRS) offers a novel, non-destructive technique allowing for rapid and low-cost soil analyses. The work for this thesis comprised two aspects of NIRS analysis: its application in the laboratory as well as in the field on-line. Although laboratory NIRS is an established method, there are no standard measurement procedures simplifying the comparability of spectral data from different NIR-devices and spectra collected over time from the same device. Therefore, the laboratory application of NIRS was investigated with the aim to optimise soil sample preparation and measurement in order to give recommendations for a standard measurement protocol. Furthermore, the on-line field application of NIRS is a relatively new method, and thus there is still a need for an evaluation of the NIR-system, manufactured by the North American company VERIS Technologies Inc., used in this study. The field application of NIRS was examined via a comparison between horizontal measurements with a shank and vertical measurements with a probe. Further investigations were carried out to test the accuracy and reproducibility of the horizontal mapping. All measurements were used to map and characterise agricultural soils in Northern Germany, with the main focus on the calibration and prediction of SOC and total nitrogen (N) concentrations and SOC stocks.

Sample-grinding and drying, as important sample preparation parameters in the laboratory, significantly decreased the calibration error for SOC when ground and oven-dried samples were used, whereas these influences were smaller for N (grinding) or had hardly any effect on calibrations (drying). A factorial analysis (ANOVA), carried out on the calibration errors for drying, ring cup and grinding, showed the biggest influence on the calibration error by grinding with 35% for SOC and 28% for N. The effect of varying the laboratory temperatures (20, 24 and 28°C), was small when oven-dried samples were used. Additionally, the reproducibility of predicted SOC and N values out of repeated NIR measurements was satisfactory for oven-dried soil samples.

Oven-dried and ground samples were therefore recommended for SOC and N calibrations, because the reproducibility within nine months was high, and particularly oven-dried samples were less sensitive to variations in laboratory temperature.

Field-based measurements led to successful SOC and N calibration results when horizontally scanned spectra were cut into the NIR region and pretreated with a standard normal variate transformation (SNV). Vertically scanned soil led to successful and excellent SOC and N calibration results of two studied fields when the reflectance or first derivative spectra of the visible and near-infrared (VIS-NIR) region were used. Universal calibrations including the vertically scanned soil with a depth of 0–70 cm always led to the best calibration results for SNV-treated spectra. Moreover, fourteen different SOC stock models were computed differing in SOC assessment and in SOC stock and BD calculation. The estimated SOC stocks differed for all models that were based on laboratory or field measurements as well as for all models including shank or probe measurements. The errors of SOC stock models were mostly lower for SOC stocks generated out of shank measurements than errors estimated via probe investigations. Furthermore, slightly less working time and costs were required when shank-based SOC stock estimations were employed, but the time difference compared to probe-based SOC stock assessments was minor.

For the reproducibility study, shank measurements were repeated after two years. It is advisable to collect and analyse reference samples for each measurement campaign, since a repeated usage of SOC data led to poorer SOC calibration results. Repeated measurements with the NIR shank led to similar SOC calibration results. Calibrations in which predominantly unmodified reflectance spectra were processed were rated as good to excellent. Moreover, SOC concentrations and SOC stocks were primarily larger when they were derived from on-line measurements than those computed out of conventionally sampled and analysed soil samples, which was attributed to the higher variation and larger range of SOC recorded via the on-line measurement. However, Kriging revealed only small differences in the distribution of predicted SOC concentrations when field measurements were repeated. The variation of SOC concentrations and SOC stocks was mostly similar for repeated measurements but the variation was slightly larger for repeated measurements in the laboratory. Furthermore, the minimum detectable difference (MDD) of SOC concentrations and SOC stocks was calculated to form further conclusions on the reproducibility of the measurement method. The relative MDDs were similar or differed slightly when the field measurements were repeated. On the basis of repeated laboratory measurements, the difference of the relative MDDs was slightly larger.

Future studies should contribute to the completion of a standard measurement protocol for laboratory NIR analyses of soil samples and to the improvement of the on-line VERIS system as a basis for efficient soil monitoring and precision agriculture.

V Zusammenfassung

Der organische Bodenkohlenstoff (OBK) spielt für die Erhaltung der Bodenqualität eine wichtige Rolle. Da er räumlich und zeitlich veränderlich ist, ist es erforderlich, den Gehalt des OBK zu messen und dessen Veränderungen festzustellen, um eine nachhaltige Bodenbewirtschaftung gewährleisten zu können. Der OBK kann mit konventionellen Laboranalysemethoden bestimmt werden, jedoch kann die Aufbereitung und Messung zahlreicher Bodenproben sehr teuer sein. Mit der Nahinfrarotspektroskopie (NIRS) ist eine zerstörungsfreie Messtechnik gegeben, mit der Bodenanalysen zeiteffektiv und kostengünstig durchgeführt werden können. Mit der vorliegenden Arbeit werden zwei Aspekte der NIRS behandelt: die Anwendung im Labor und auch die Anwendung im Feld *on-line*. Auch wenn die NIRS eine im Labor bewährte Messmethode darstellt, gibt es nach wie vor kein standardisiertes Messverfahren, das die Vergleichbarkeit von Bodenspektren verschiedener NIR-Geräte bzw. die Vergleichbarkeit von Spektren, welche von einem Gerät über einen längeren Zeitraum aufgenommenen wurden, vereinfacht. Daher wurde in dieser Arbeit eine Optimierung der Probenvorbereitung und -messung angestrebt, um Empfehlungen für ein einheitliches Labormessprotokoll geben zu können. Die NIRS Anwendung im Feld ist eine recht neue Methode, so dass das speziell in dieser Arbeit verwendete NIR-System der nordamerikanischen Firma VERIS Technologies Inc. einer weiteren Bewertung bedarf. Hierbei wurden horizontale NIR-Messungen mit einem Pflug und vertikale NIR-Messungen mit einem Bohrgerät durchgeführt und die aufgenommen Bodendaten miteinander verglichen. Ebenso wurde die Genauigkeit und Reproduzierbarkeit der mit dem Pflug aufgenommen Bodendaten untersucht. Alle hier durchgeführten Messungen dienten der Datenaufnahme und Charakterisierung von Ackerböden Norddeutschlands mit dem Ziel, OBK- und Gesamtstickstoffgehalte sowie OBK-Vorräte zu kalibrieren und zu schätzen.

Ergebnisse aus dem Labor ergaben, dass der Zermahlungsgrad und das Trocknen wichtige Parameter für die Aufbereitung von Bodenproben darstellen, da der Kalibrationsfehler von OBK bei Verwendung gemahlener und ofengetrockneter Proben signifikant kleiner wurde. Die Einflüsse der vorgenannten Bodenaufbereitungen waren für den Stickstoff (N) kleiner als für OBK oder zeigten nur einen geringen Einfluss auf die Kalibration (Trocknung). Eine durchgeführte Faktorenanalyse (ANOVA) aller Kalibrationsfehler zu Trocknung, Zermahlungsgrad und Bechergröße zeigte den größten Einfluss auf den Kalibrationsfehler für den Zermahlungsgrad mit 35% für OBK und 28% für N. Der Einfluss veränderlicher Labortemperaturen (20, 24 und 28°C) zeigte lediglich einen kleinen Effekt, wenn

ofengetrocknete Bodenproben verwendet wurden. Wurden diese Proben wiederholt gemessen, ergab sich eine zufriedenstellende Reproduzierbarkeit für geschätzte OBK- und N-Werte. Aufgrund der hohen Reproduzierbarkeit können somit gemahlene und getrocknete Proben für OBK- und N-Kalibrationen empfohlen werden, wobei insbesondere getrocknete Bodenproben eine nur geringe Sensitivität gegenüber variierenden Labortemperaturen aufwiesen.

Horizontale Feldmessungen mit dem NIR-Pflug führten zu erfolgreichen Kalibrationsergebnissen von OBK und N, und zwar bei Verwendung von Feldspektren, welche das NIR-Spektrum umfassten und mit der *standard normal variate* Transformation (SNV) mathematisch vorbehandelt wurden. Vertikale Feldmessungen mit dem NIR-Bohrgerät ergaben vornehmlich erfolgreiche bis exzellente Kalibrationsergebnisse, wenn Reflexionsspektren oder Spektren der ersten Ableitung des gesamten sichtbaren und Nahinfrarot-Spektralbereiches (VIS-NIR) für die Kalibration verwendet wurden. Für eine Kalibration von Bodenproben aus einer Tiefe von 0-70 cm führten die mit SNV behandelten Feldspektren zu durchgehend besten Kalibrationsergebnissen. Darüber hinaus wurden 14 verschiedene Modelle für OBK-Vorräte berechnet, die sich in der Erfassung von OBK sowie in der Berechnung der Lagerungsdichte und der OBK-Vorräte unterschieden. Die ermittelten Kohlenstoffvorräte variierten für alle Modelle, unabhängig davon, ob sie aus Feld- oder Labordaten berechnet wurden und auch unabhängig davon, ob die Feldmessungen mit dem Pflug oder dem Bohrgerät durchgeführt wurden. Die Fehler dieser Modelle waren zumeist kleiner, wenn die Modelle mit Daten aus Pflugmessungen generiert wurden. Modelle, welche Daten aus Bohrgerätmessungen enthielten, wiesen meist größere Fehler auf. Für die Bestimmung von Kohlenstoffvorräten aus Pflugmessungen wurden darüber hinaus weniger Arbeitszeit und Kosten benötigt, wobei der Unterschied zu aus Bohrgerätmessungen generierten Modellen klein war. Um die Reproduzierbarkeit zu untersuchen, wurden die Feldmessungen mit dem Pflug nach etwa zwei Jahren wiederholt. Es ist empfehlenswert, für jede Feldmessung eine eigene Bodenprobenentnahme mit eigener Laboranalyse durchzuführen, da eine wiederholte Verwendung von OBK-Daten einer früheren Messkampagne zu schlechteren Kalibrationsergebnissen führte. Die aus wiederholten Pflugmessungen berechneten Kalibrationen für OBK waren in ihrem Ergebnis vergleichbar. Gute bis exzellente Kalibrationensergebnisse konnten zumeist unter Verwendung unveränderter Reflexionsspektren erreicht werden. Die aus Feldmessungen berechneten OBK-Gehalte und deren Vorräte wiesen im Wesentlichen höhere Werte auf, als die aus den Labormessungen berechneten. Man ging hierbei davon aus, dass die Feldmessungen eine größere Variation und einen größeren Konzentrationsbereich von OBK erfasst haben. Ein

Kriging ergab nur kleine Unterschiede in der Verteilung der geschätzten OBK-Konzentrationen, wenn die Feldmessungen wiederholt wurden. Ebenso waren die Variationskoeffizienten von OBK und deren Vorräten aus Wiederholungsmessungen im Feld fast durchgängig gleich. Der Unterschied zu den aus dem Labor berechneten Variationskoeffizienten war geringfügig größer.

Weiterhin wurde die minimal detektierbare Differenz (MDD) von OBK- und deren Vorräten berechnet, um hieraus weitere Rückschlüsse auf die Reproduzierbarkeit der Messungen zu ziehen. Die Berechnung der relativen MDDs für wiederholte Feldmessungen ergaben gleiche Werte, oder solche, die sich nur minimal voneinander unterschieden. Der Unterschied der relativen MDDs aus Wiederholungsmessungen im Labor war hingegen geringfügig größer.

Zukünftige Studien sollten an der Fertigstellung eines standardisierten NIR-Labormessprotokolls für Bodenproben mitwirken und auch das VERIS-System als Grundlage für effizientes Bodenmonitoring und *precision agriculture* weiter verbessern.

1 Introduction

With a history of approximately 150 years of industrial activity, the developed countries are largely responsible for the high levels of greenhouse gas emissions. Therefore, the United Nations Framework Convention on Climate Change (UNFCCC) and its Kyoto Protocol have established important guidelines to solve international environmental problems by setting binding emission reduction targets. The countries involved are to reduce their emissions by at least 18% below 1990 levels in the current second commitment period from 2013 to 2020. The greenhouse gases of primary concern are carbon dioxide (CO_2), methane and nitrous oxide. Their concentrations have steadily increased over the time period between 1850 and 2005 (Lal, 2008) and are still increasing at present. The influence (proportion) from the greenhouse effects of CO_2 and methane is, for both, around 20% and of nitrous oxides is 5%. CO_2 is the most human-contributed greenhouse gas, and it is one part of the global carbon cycle characterised by incomes and losses of carbon between five carbon reservoirs (oceanic, geologic, pedologic, atmospheric and biologic). The present study addresses the carbon stored in agricultural soils, which represent the pedologic pool. Generally, the soil system has been left imbalanced by land use change and tropical deforestation in the past as well as in the present, and huge amounts of emissions have decreased the soil carbon pool (Lal, 2004). The loss of the SOC pool, as one part of the soil carbon pool, from agricultural soils may be as much as 30-60 Mg ha^{-1}, depending on climate, land use and management systems (Lal, 2001). According to the Kyoto Protocol, mitigating the increasing atmospheric CO_2 concentrations by carbon sequestration in terrestrial ecosystems is a promising low cost option. Here, agricultural croplands have great potential for sequestering atmospheric carbon. Numerous studies have been conducted to analyse the effects of management on SOC sequestration because of the increasing interest in providing a sink for atmospheric carbon. When carbon is sequestered, it remains in the soil as long as best management practices and restorative land use are followed.

Examples of strategies to increase the soil carbon pool are reducing tillage intensity, changing crop rotations, improving fertilizer management and using winter cover crops.

In order to be able to give suggestions for best management strategies, the monitoring of SOC changes in agricultural soils is essential. But it is still a challenge to quantify these changes accurately due to their large variability. As Nash et al. (2011) have pointed out, a general lack of availability of the required data impedes the realisation of a potential automated compliance assessment according to common agricultural management standards. Long-term data collection is crucial for explaining the spatial and temporal changes of SOC

concentrations and also stocks. For this purpose, intensive and reliable soil mapping methods are required.

The current and future trends of technologies for monitoring soil carbon sequestration are developing in the direction of cost- and time-effective measurement methods.

There is widespread interest in NIRS to measure soil properties because the technique is rapid, relatively inexpensive and requires minimal sample preparation. The work for this thesis addressed two aspects of NIR analysis: laboratory as well as *in-situ* field application.

The laboratory approach of NIR measurements was first developed for the analysis of moisture in grains (Bengera and Norris, 1968). From the 1980s on, NIRS has been successfully used to predict SOC and the total N content of soils (Krishnan et al., 1980; Dalal and Henry, 1986; Sudduth and Hummel, 1993; Ben-Dor and Banin, 1995b; Ludwig et al., 2002; He et al., 2005; Brunet et al., 2007; Nocita et al., 2011). However, soil chemical components such as SOC and N are of low concentrations, and analysis is exacerbated by a heterogeneous sample matrix. Although NIRS is an established method for laboratory analysis, the calibration results are conflicting throughout the literature concerning soil sample preparation and measurement. Several methodological aspects, such as sample-grinding and drying, have been identified as important factors, but the results on the best methods remain inconsistent.

However, spectral measurements carried out in the laboratory using a limited number of soil samples are often not sufficient for a detailed description of soil heterogeneity at the field scale. Therefore researches have focussed on developing sensors for in-field measurements of soil properties. Here, successful on-line measurements would be of great benefit for management zones in precision agriculture, aiming at better land management and the reduction of the amount of inputs applied into the environment (Malhi et al., 2001). Furthermore, choosing an appropriate management practice for crop production helps farmers to increase their profit. In recent years, soil sensing technologies such as the use of ion selective electrodes to measure soil pH (Sethuramasamyraja et al., 2008), ground penetrating radar for measurements of soil water content (Pettinelli et al., 2007) and cameras to measure soil colour for estimating soil carbon content (Rossel et al., 2008) have been reported as being useful for various in-field applications in precision agriculture. However, the VIS-NIR spectroscopy has proven to be the most capable soil sensing technology for on-line measurements of various soil chemical and physical properties at the field scale (Shibusawa et al., 2000a; Mouazen et al., 2007a; Marin-Gonzalez et al., 2013). The tractor-mounted system, such as the one constructed by Christy (2003), is one example of VIS-NIR sensors. Thus far,

there have been only a few studies conducted using the tractor-mounted system, so various investigations necessary for system evaluation are missing.

Overall, the laboratory aspect of this work deals with the optimisation of NIRS, which is a prerequisite to establishing NIRS as a soil analytical method. The objectives are as follows:

- to comprehensively investigate the impact of sample preparation, such as sample-grinding and drying, on NIR calibration quality and prediction errors for SOC and total N concentration;
- to examine the effects of different measurement parameters, such as varying laboratory temperature and repeated NIR scans on NIR calibration results and reproducibility and
- to give recommendations for a standard measurement protocol for soil sample NIR scans.

The field aspect of this work deals with a method comparison between horizontal and vertical soil data acquisitions via on-line NIRS. The objectives are as follows:

- to compare horizontal and vertical field NIR measurement methods with respect to NIR calibration error for SOC and total N concentration;
- to find the best model out of different SOC stock model options based on lowest time expenditure and highest accuracy and
- to compare both methods in consideration of time and costs required for field-scale SOC stock assessments.

Moreover, the accuracy and reproducibility of the horizontal on-line NIR mapping of topsoils were investigated. Therefore, repeated measurement campaigns were carried out within short time intervals. The objectives are as follows:

- to find out whether repeated measurements lead to comparable field-scale SOC concentrations and stocks when they are carried out on-line in the field or in the laboratory via direct chemical analysis;
- to visualise the spatial SOC distribution on two agricultural fields and to identify similarities and differences of repeated measurement campaigns and
- to determine minimum detectable differences (MDD) of field-scale SOC stocks for different field inventory methods.

2 State of the art

2.1 Outline of the history of soil mapping

Very early soil knowledge, as gained by the Greeks, was mainly based on observations of nature without any experiments to test the theories. The first agricultural activities came from the modern village of Jarmo, Iraq, dating back to 11000 years before present, and it was mostly a trial and error approach to decide where to farm (Troeh, 2004). In the 16th century and the beginning of the Renaissance in Europe, science and scientific thinking started to develop. During the 19th century, the soil profile, as a vertical section of the soil from the ground surface to the underlying parent rock, developed as a major soil science concept. The first German scientific works on soils began with studies from Senft (1857), Fallou (1862) and Orth (1877), who used soil profiles as an information basis to create soil maps based on soil texture and humus content. The soil maps were mainly needed for tax assessment and for the selection of the type of land management. Around 1930, J. Görbing developed the spade diagnosis to have a simple-to-use and cost-effective field method for the evaluation of the significant ecological structure parameters of the soils under agricultural management. This estimation method had a strongly descriptive character and was later extended by Hampl (1994) and Beste (2002) as a scientific method for soil evaluation for the upper 30 cm soil depth. Even though most studies and inventories on SOC are confined to the 30 cm soil depth (Beste, 2002; IPCC, 2003; Smith et al., 2005) the amount of SOC stored below 30 cm is relevant in many ecosystems (Batjes, 1996; Jobbagy and Jackson, 2000).

Today, the soil profile of the 100 cm soil depth is an established method to determine soil quality (spatial determination of soil attributes under different management practices) such as by the German Forest Soil Inventory (BZE I: 1987–1993, BZE II: 2006–2008), the German Agricultural Soil Inventory (2011–present) and the soil fertility appraisal. Since SOC is a highly spatially variable as influenced by differences in soil types, texture and soil redistribution (VandenBygaart, 2006), the mentioned soil inventories added drilling cores around each soil profile to derive the site variability at the intersection points of an 8 x 8 km grid. The eight positions of the drillings points were chosen within a 10 m radius around the soil profile at all cardinal or half-cardinal points. However, the spatial variability of SOC is present at different scales (Bird et al., 2002), and it is becoming increasingly more important to improve the spatial resolution of maps as a fundamental information layer for studying ecological processes and to overcome land degradation. Regarding the field scale, within-field variability, including spatial and temporal aspects, is the most important requirement for the successful implementation of precision agriculture (Srinivasan et al., 2006). The conventional

characterisation of within-field spatial variability usually comprises manual sampling, sample pre-treatment and laboratory chemical and physical analyses. The need for soil sampling has always had a particularly economic basis (Peck and Soltanpour, 1990). And since at least the 1920s, the number of soil samples that represent field variability has been an important topic of discussion. Over the past decades, a range of soil sampling approaches has been generated for mapping natural resources. The economic goal has always been to sample representatively for the area under study while keeping the samplings costs low (Sawyer, 1994). However, conventional soil analysis can be very expensive and time consuming, depending on a field´s variability (Kitchen et al., 2008). Therefore, new measurement methods are needed to replace the conventional reference methods for providing intensive information about soils in a timely manner. Recent advances in technology offer large amounts of data with high resolution acquired through proximal and remote soil sensing techniques. While remote sensing includes airborne and satellite sensors, proximal soil sensing is used to measure soil properties with field sensors. Here, NIRS is one of the most promising measurement techniques for fast, cost-effective and environment friendly data collection of within-field variability.

2.2 NIRS application in soil science

NIRS has been widely used in agriculture for decades. The application of NIRS in soils started in the 1990s. Up until 2011, over 200 publications have been produced, that cover the prediction of chemical, physical and biological soil attributes, with 50% of these in the last three years alone (Bellon-Maurel and McBratney, 2011). Since then, many more investigations have been carried out as well. NIRS has been found to be useful in measuring soil properties because the technique is accurate, reliable, rapid, less expensive and non-destructive, and it consumes no reagents and can measure several soil components simultaneously (Chang and Laird, 2002; Chang et al., 2005; Huang et al., 2007; Christy, 2008; Munoz and Kravchenko, 2011).

NIR and VIS spectroscopy, moreover, have an advantage over some of the conventional techniques of soil analysis in being more efficient when a large number of soil samples and analyses are needed. NIRS can lower laboratory costs by at least an estimated 80% (Nduwamungu et al., 2009a). Furthermore, it has strong potential for detecting soil carbon (Viscarra Rossel et al., 2006). The spectra of complex organic materials are not directly informative since they are an addition of elementary absorptions corresponding to several bonds. NIRS is an indirect analytical method based on the development of empirical models that can predict the concentration of soil components from spectral data. Therefore, four steps

are needed for a successful prediction; these are i) the NIR measurement; ii) the pretreatment of spectral data; iii) the calibration, in which regression models are developed via the use of a subset of soil samples and iv) the validation, where the validity of the regression models is tested via the remaining soil samples for the prediction of component concentrations.
Generally, infrared radiation is the region of the electromagnetic spectrum between the visible (VIS) and the microwave wavelength (Figure 1) and is divided into near- (NIR), mid- (MIR) and far- (FIR) infrared parts, depending on their relation to the visible spectrum.

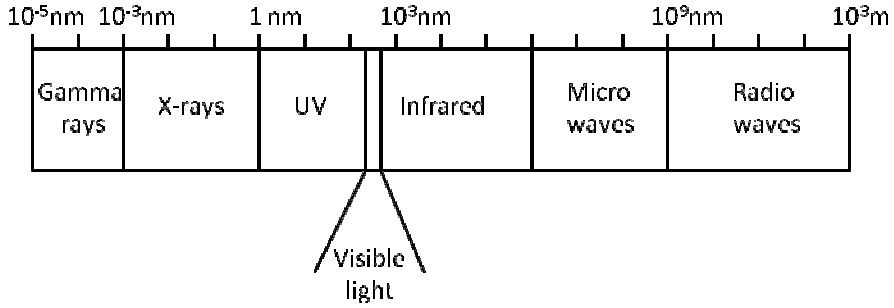

Figure 1: Overview of the electromagnetic spectrum.

Infrared spectroscopy involves the use of light to probe the vibrational behaviour of molecular systems to identify and study chemicals. Overall, absorption spectra are a result of physical light interactions, such as particle size and soil structure, and chemical light interactions (molecular bonds absorb energy). NIR spectra are primarily composed of the stretching and bending of the CH, OH and NH groups and contain information from organics and the hydroxyls from organics, amongst others. The weak and broad signals come from the vibrations from overtones and combination bands. In contrast, the MIR region is dominated by intense vibrational fundamentals with the additional vibrations of the CO, CN and CC groups that deliver information on minerals. However, MIR spectroscopy is not as established as NIRS for the prediction of different soil constituents such as SOC and N (Viscarra Rossel et al., 2006; Michel et al., 2009), probably due to the weaker absorbance of moisture in NIR that allows the use of field-moist samples (Christy, 2008) and NIRS' lower demand regarding sample preparation.

2.3 Laboratory and field-based NIR measurements

Over the past two decades, three particular scales of NIR spectroscopic methods have been applied for SOC assessment: laboratory spectroscopy, field spectroscopy and imaging spectroscopy. Imaging spectroscopy has often been carried out for regional studies via airborne measurements. Even though the airborne mapping of SOC has made considerable progress, it is costly and highly weather-dependant (Hbirkou et al., 2012). Generally, soil scientists use various conventional laboratory analytical methods for determining soil chemical and physical properties, and NIRS is among these methods. As Viscarra Rossel et al. (2006) stated, the benefit of infrared spectroscopy is that it is easier to handle compared to most conventional techniques. Specifically, NIRS has the advantage of being adaptable for on-line field use. Here, it meets the need for large amounts of high quality data recorded with low costs. Therefore NIRS is considered to be a possible alternative to conventional methods of soil analysis (Munoz and Kravchenko, 2011).

Overall, the capability of NIR laboratory spectroscopy for the determination of soil properties has frequently been confirmed in the literature. This technique has been used to estimate various chemical soil properties, such as the SOC and total N content in agricultural soils (Ludwig et al., 2002; Moron and Cozzolino, 2002; Reeves et al., 2002; He et al., 2005; Deng et al., 2013; Gao et al., 2014; Peng et al., 2014) and macrominerals and micronutrients in soils (Chang et al., 2001; Dunn et al., 2002; Malley et al., 2002; Viscarra Rossel et al., 2006; Viscarra Rossel et al., 2009; Szalai et al., 2013) as well as various physical soil properties (Chang et al., 2001; Dunn et al., 2002; Brunet et al., 2007; Nduwamungu et al., 2009b; Wight et al., 2016).

However, soil NIR spectroscopy is currently operating at the edge of technical feasibility, since target soil parameters are often present at concentrations far below 5% and often in the trace range. The matrix of soils has various chemical and physical properties that interfere with the measurements. Quartz in soils, for example, has a large reflectance in the NIR region (Hunt and Salisbury, 1970) and can show up as an interfering component (Viscarra Rossel et al., 2006). Several factors affect the NIR spectra of soil samples, including sample pretreatment, sample presentations and laboratory as well as spectrometer conditions. Certain factors have been investigated (Nduwamungu et al., 2009a), but some results remain inconsistent and conflicting, such as the influence of sample preparation.

Up to now, most studies have investigated the effects of sample size and soil properties (Chang and Laird, 2002; He et al., 2005; Cozzolino and Moron, 2006; Wetterlind and Stenberg, 2010). Fewer studies cover the effects of sample preparation, sample measurement

and repeated measures (Fystro, 2002; Barthes et al., 2006; Brunet et al., 2007). Two studies recommended the analysis of ground soil samples for calibration (Barthes et al., 2006; Brunet et al., 2007), but the accurate NIRS prediction of SOC content using unground samples was also reported by Chang et al. (2001), Moron and Cozzolino (2004) and Shepherd and Walsh (2007). Moreover, some reports have even suggested that grinding the samples did not improve prediction accuracy. However, no clear benefit was seen for different types of sample drying, the number of repeated measurements (Barthes et al., 2006) or for different sample cups for calibration (Fystro, 2002).

Thus far, there has been no comprehensive study that has investigated the effects of soil sample preparation techniques (grinding and drying), NIR scanning protocols (replicates, size of scanned sample surface) and laboratory and spectrometer conditions (temperature, air humidity) to elucidate the major factors that influence the quality of soil NIR analysis in the laboratory.

While the spatial and temporal variability of soil properties at field scales makes accurate quantification and assessment more difficult, NIR field sensors are attracting increasing attention because of their potential to record, *in-situ*, huge amounts of data with high spatial resolution. This method is relatively new and has been carried out both statically and on-line over the last 15 years. In-field measurements of soil reflectance are generally influenced by conditions such as soil moisture, structure, coarse organic residues and the contamination of the sensor by dust (Stenberg, 2010; Gubler, 2011). Though these conditions are not favourable to characterise soil properties via NIRS, publications report promising results for the SOC and total N content assessment of soils (Ben-Dor et al., 2008; Viscarra Rossel et al., 2009; Hedley et al., 2010; Knadel et al., 2011; Kuang and Mouazen, 2013).

Various static soil sensors and techniques have been developed – for example, the use of electrical resistivity to measure soil electrical conductivity (Sudduth et al., 2001), magnetic susceptibility to measure soil contamination by heavy metals (Jordanova et al., 2008), cameras to measure soil colour for estimating soil carbon content (Viscarra Rossel et al., 2008) and many more. Predicting soil C via static NIR measurements has been successfully carried out by Ben-Dor et al. (2008) and Viscarra Rossel et al. (2009) via the use of portable spectrophotometers. However, rapid NIR measurements in the field, which are performed on-line, may generate an even greater amount of data needed for the large spatial coverage of soil C. Here, successful predictions of soil C via on-line NIRS have been performed since 2009.

There are three on-line sensors currently available (Kuang et al., 2012) that measure the soil properties of the topsoil, predominantly at a depth of around 7–10 cm. All three sensors are

mounted onto a tractor that pulls the system through the soil. The first on-line NIR sensor was constructed by Shibusawa et al. (2000a), followed by the VIS-NIR shank from Christy (2008) and the NIR spectrophotometer from Mouazen (2009), the latter of which is a simpler design than that of Shibusawa. Up to now, pH as well as soil organic matter have been predicted via the use of all three sensors. Moisture content, electrical conductivity and soil temperature could be partly predicted by these. The VIS-NIR shank from Christy (2008) is commercially sold via the VERIS Technologies Incorporation in Kansas, USA. Their system has been available since 2006 and the probe, as its pendant, since 2007. Successful NIR data logging can be found in a few papers since 2007 that solely used the application of the VERIS shank for organic matter (OM) and carbon content assessment (Huang et al., 2007; Christy, 2008; Knadel et al., 2011; Munoz and Kravchenko, 2011), as well as for electrical conductivity (EC) and pH analyses (Adamchuk et al., 2011). The mentioned studies performed horizontal on-line NIR measurements in soils with different SOC contents, textures and topographies. There has been no study taking into account vertical on-line NIR measurements via the VERIS probe for predicting the SOC content of agricultural subsoils. Comparing topsoil and subsoil information as a two- and three-dimensional approach, based on NIRS, has also not been studied thus far.

Since the carbon reservoir is not permanent but rather enters and leaves the soil in the form of a dynamic equilibrium, there is a need for more than one ‘snapshot’ of the soil’s characteristic: Repeated measurements are necessary in order to detect changes in SOC contents.

Generally, the temporal change of SOC stocks is difficult to detect, because changes are typically small in comparison to the total amount of SOC (Schoening et al., 2006). In addition, the high spatial variability of SOC stocks impedes field-based measurements, because a very large sample number is needed to detect changes. As Schrumpf et al. (2011) found, around 100 soil samples are needed to detect a change in SOC within 10 years. Conen (2005) estimated that a SOC stock change between 120 and 2480 g m^{-2} can be detected using a sample size of 100. And Smith (2004) additionally concluded that a minimum of 10 to 15 years must pass before an SOC stock change can be detected. However, the underlying methods of the published studies on SOC stock change detection and SOC stock changes vary considerably; thus, there is a need for harmonised measurements generating comparable data.

3 Materials and Methods

This chapter is divided into three sections: it covers the laboratory aspect of NIRS, called 'laboratory-based NIR measurements', and the field aspect of NIRS split into the themes 'field-based NIR measurements' and 'field reproducibility study'. For overview purposes, this structure has been maintained in the 'results', 'discussion' and 'conclusions' sections.

3.1 Laboratory-based NIR measurements

3.1.1 Sample preparation

97 soil samples from agricultural soils (0-50 cm depth) were used that covered a broad range of soil types from Northern Germany. The samples were air-dried and sieved to pass a 2 mm screen. Two subsamples of 30 g per sample (100 g) were taken for further treatments. One subsample was crushed for 1.2 minutes in a pebble mill (Retsch GmbH, Haan, Germany, RM200), called "crushed sample", and one subsample was ground for 1.2 minutes with a frequency of 30 s^{-1} (Retsch GmbH, Haan, Germany, MM400). Material of grinding jar and grinding balls was zirconium oxide. The SOC content of the sample set ranged between 0.23 and 3.81% and the total N content between 0.05 and 0.37% (Figure 2).

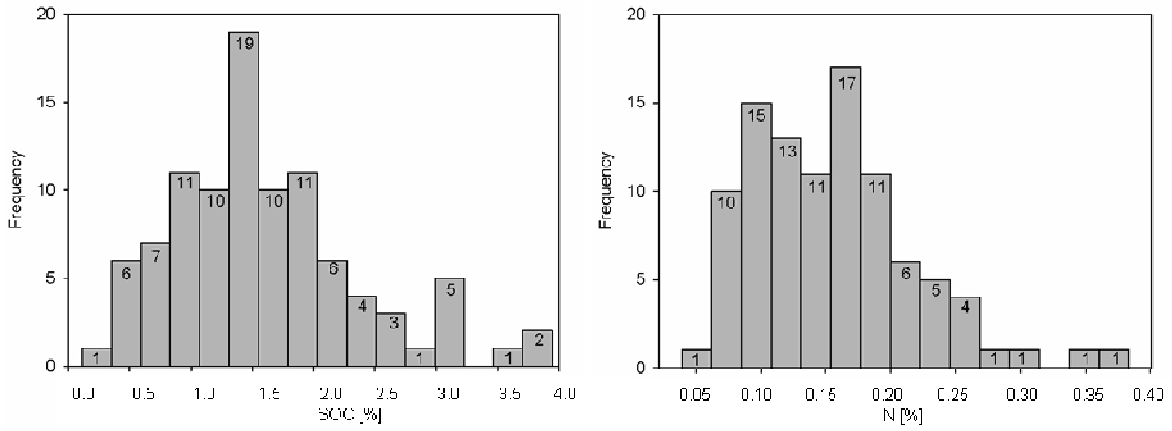

Figure 2: SOC and total N content of soil samples and their frequency distribution.

3.1.2 Reference analyses for SOC and N

All ground soil samples were dried at 40°C until constant weight. Total C and total N content were determined via hot combustion (Vario Max, Elementar Analysensysteme GmbH, Hanau, Germany). Soil inorganic carbon (SIC) content was determined after ignition at 450°C for 16 hours in a muffle kiln. SOC content was determined as the difference between total C and SIC content. Both concentrations were corrected for free water in the air-dried samples.

3.1.3 NIR spectra acquisition

Reflectance spectra were measured between 910-2631 nm at 1 nm intervals with a Multi Purpose Fourier Transform (FT) NIR spectrometer (Bruker Optik GmbH, Ettlingen, Germany). For each scanning the ring cup was filled with about 20 g of soil sample. All soil samples in Experiments 2 to 4 were measured twice (for obvious outlier spectra also three times) and the spectra were averaged before calibration.

3.1.4 Sample treatments

For investigating effects of different sample treatments on NIR calibration quality and prediction results for SOC and N a total of 34 NIR spectrum sets were recorded (Table 1).

Repeated NIR scans: Experiment 1

The first effect investigated the influence of repeated measures. Sieved samples were scanned four times. Out of this spectrum set, spectra were taken randomly for analyzing replication effects for one, two, three and four measurements. The random choice yielded in 1000 simulations for PLS (partial least squares) analysis for all four tested replications.

Spectra were averaged for PLS regressions before cross validation. The results were compared in particular with those for ground samples. For spectroscopic measurements a ring cup was independently refilled with the soil sample on target for each spectrum acquisition.

Sample preparation: Experiment 2

The second experiment analysed the factorial effect of twelve different sample pretreatments concerning different grinding levels in combination with different sizes of ring cups and degrees of drying (see Table 1). Oven-dried samples were put in a desiccator for cooling until measured. Two ring cups with different sizes of scanning area for measurements were used: 19.6 cm^2 was scanned via the big cup and 6.8 cm^2 via the small one. Quartz glass was the bottom material for both ring cups. A three-way ANOVA was performed to see whether the RMSECV was affected by grinding, ring cup size and drying. The sum of squares of each treatment was converted into percent to more clearly see the proportion of effects.

Table 1: Overview on tested influences on NIRS measurements and sample treatments

Experiment No.	Investigated effect	Treatment	Measurement conditions	Target values	No. of spectrum sets	No. of soil samples
1	repeated measures	one, two, three, four measurements	sieved, air-dried	RMSECV [a], R^2 [b], RPIQ [c]	6	94
2	grinding level, drying condition, size of ring cup	sieved, crushed, ground air-dried and oven-dried measured in a small and a big ring cup	factorial experiment	RMSECV, R^2, RPIQ	12	95
3	soil temperature	20, 24, 28°C	ground samples, air-dried and oven-dried	RMSECV, R^2, RPIQ, RPD [d], RMSEP [e], r^2 [f]	6	97
4	fluctuating laboratory humidity and temperature (1), reproducibility (2)	Directly after drying and after one, two, five weeks and 9 months later, also after drying	ground (1,2) and sieved samples (2) air-dried (1,2) and oven-dried (1,2)	RMSECV, R^2, RPIQ, RPD, RMSEP, r^2	10	96

[a]root mean square error of cross validation, [b]coefficient of determination of calibration
[c]ratio of performance of IQ, [d]ratio of prediction to deviation,
[e]root mean square error of prediction, [f]coefficient of determination of prediction
Fluctuating lab conditions and reproducibility: Experiment 4

Sample temperature: Experiment 3

The third experiment investigated the effect of different soil temperatures on NIR measurements. Investigations were carried out in a climate chamber with a temperature controlling system. Two thermometers were used to check the room and soil temperature. Spectrum acquisition could start after the target temperature was reached. All samples were analysed in oven-dried and air-dried condition to find out if differences in NIR results might accompany different degrees in drying. As a basis, soils with 20°C served for calibration, and soils with 24°C and 28°C were tested for prediction. A two-way analysis of variance was carried out to see any effect on RMSEP by varying temperature.

The fourth experiment was developed to observe any influence of fluctuating lab conditions (not controlled by temperature and humidity) on not standardised dried soils. Oven-dried samples were put in an open aluminum cup with direct contact to the air. For each measurement, the samples were transferred from the aluminum cup to the ring cup. Double

scans of the 97 samples were carried out within a maximum of three days for each treatment. They were analysed directly after cooling and then again in certain time intervals. Air-dried samples served for calibration and all other spectrum sets for prediction. Additionally the reproducibility of calibration results was tested on three sample pretreatments: on air-dried sieved, air-dried ground and oven-dried ground samples. Measurements were repeated in a time interval of nine months.

3.1.5 Pretreatment of spectral data

All spectra were converted from cm^{-1} in nm and cut into a wavelength range from 1250-2631 nm since outside this range there was hardly any distinct signal and the selected wavelength range always resulted in the lowest calibration errors. Spectra manipulation such as smoothing, first derivative, SNV and normalization did not lead to significantly better calibration results. Thus, the original reflectance spectra were taken for all further investigations.

3.1.6 Multivariate data analysis

Multivariate data analysis is used for different purposes with the objectives of data description, discrimination and classification, regression and prediction (Esbensen, 2006). In the VIS-NIR region, water absorbs over a large range of wavelengths and overlaps with other important peaks. This is why it is not possible to use absorbance at a single wavelength to predict the concentration of absorbers. The solution is to combine many different wavelengths in order to assign specific features to specific chemical components – so-called multivariate data analysis. The most used mathematical methods in soil science include Partial Least Squares Regression (PLSR) (Reeves and Zapf, 1999; Martin et al., 2002; McCarty et al., 2002; Cozzolino and Moron, 2003; Udelhoven et al., 2003; Chang et al., 2005; Sorensen and Dalsgaard, 2005; Viscarra Rossel et al., 2006; He et al., 2007; Viscarra Rossel et al., 2009), Principal Component Regression (PCR) (Chang et al., 2001; Pirie et al., 2005; He et al., 2007), Multiple Linear Regression (MLR) (Dalal and Henry, 1986; Ben-Dor and Banin, 1995a; Malley et al., 2002) and Stepwise Multiple Linear Regression (SMLR) (Shibusawa et al., 2000a; Shibusawa et al., 2000b), which are all common linear models. Non-linear models include regression trees (Brown et al., 2006) (Viscarra Rossel and Behrens, 2010), Neural Networks (NN) (Mouazen et al., 2010) and Multivariate Adaptive Regression Splines (MARS) (Shepherd and Walsh, 2002; Viscarra Rossel and Behrens, 2010), among others.

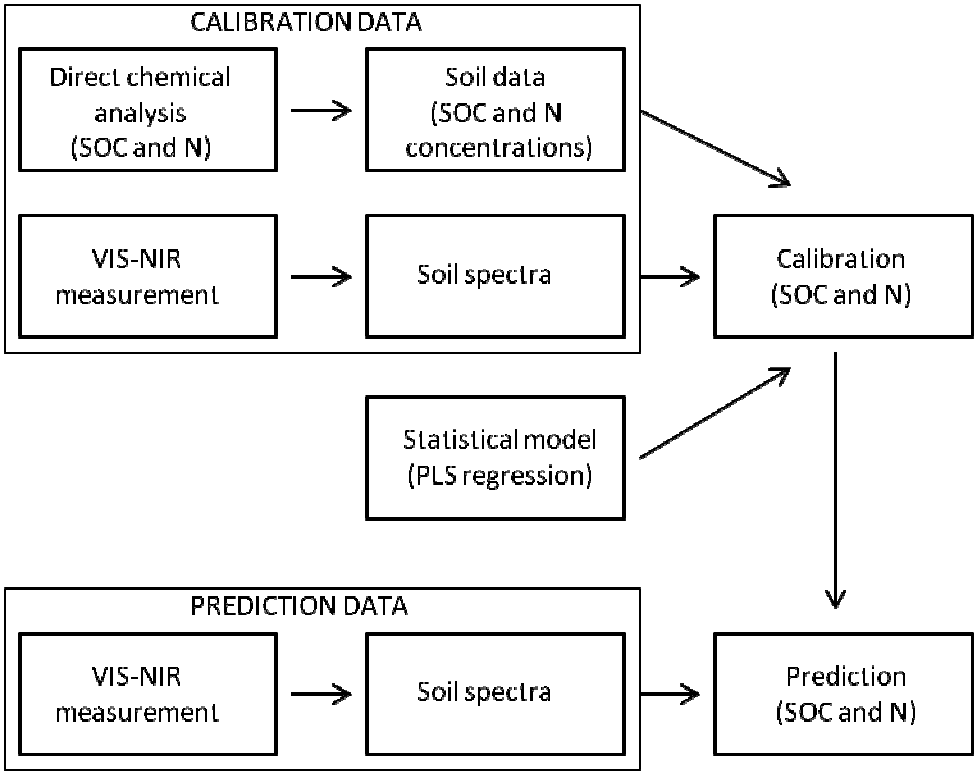

Figure 3: Overview of the information flow for a multivariate calibration of soil spectra and SOC and N concentrations as well as the corresponding statistical model that enables SOC and N prediction out of other soil spectra.

For multivariate data analysis, the NIR spectroscope requires indirect calibration: A set of calibration samples with measured soil components (SOC and N concentrations via direct chemical analysis) is measured again via NIRS. These data are used to build a mathematical model. Once calibrated, the NIR spectroscopy allows a rapid and precise analysis so that concentrations of new soil samples can be easily predicted (Figure 3). Using statistical models, Krishnan et al. (1980) established relationships between soil reflectance and soil organic matter as early as 35 years ago.

3.1.7 PLSR for calibration

The purpose of quantitative measurements via NIRS is to predict certain useful information. A mathematical formula is required in order to transform the measured data into relevant information. The PLS regression is one of the most commonly used linear regression type. It is used to explain how predictor variables (X) explain the variations in response variables (Y). PLSR models the X and Y matrices simultaneously to discover the latent variables (components) in X that best predict the components in Y (Wold et al., 1984; Martens and Naes, 1989).

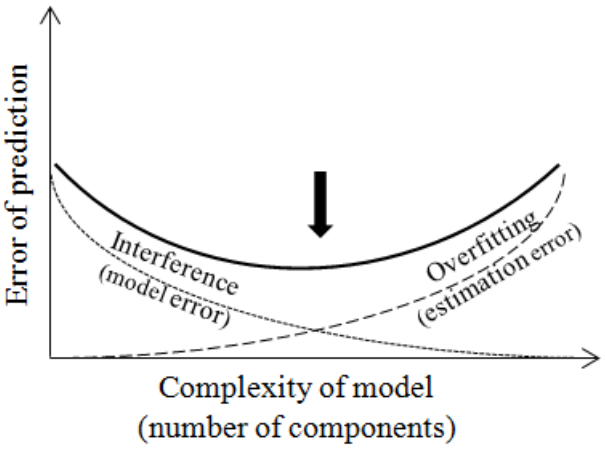

Figure 4: Selecting the right model complexity of a calibration model. Interference and overfitting have a negative effect on prediction ability. Estimation error and model error contribute to prediction error. (Modified after Martens and Naes (1989)).

In PLS regression, each component is obtained by maximising the covariance between Y and all possible linear functions of X. This regression technique reduces noise and data dimensionality and is computationally fast. The result is a set of calibration models that are based on a certain number of components. In order to select the best model for optimal prediction, the relation between the number of components and the prediction error must be examined (Figure 4). As the complexity of a model increases, meaning the number of components in the model increases, the model error decreases, since more of the spectral variability is modelled, and the systematic interference from other chemical constituents decreases. At the same time, the estimation error increases, which is due to random errors in the calibration data. Optimal prediction is obtained by balancing the two error types. Here, the RMSECV has the lowest value.

3.1.8 Calibration and determination of best models

Chemometric analyses were performed using Calibration Wizard version 1.1 (SensoLogic GmbH, Norderstedt, Germany). For the calculation of the Mahalanobis distance and to detect outliers, a PLS regression was carried out on spectral data using a leave-one-out cross-validation method. The ‘H’, ‘T’ and ‘D’ values were calculated for each spectrum to determine whether there were possible outliers. H-outliers are also called influence outliers, since they show how strong the influence of a particular spectrum is on the regression model. More precisely, they are a measure of the multidimensional distance of a spectrum from the regression line. Small sample sets are especially vulnerable to these outliers. T-outliers can be detected from a Student’s t-test carried out for each spectrum. This determines the residual

error, or how closely the reference value matches the predicted value. D-outliers can be detected from a Cook's statistic that takes into account both the T and H values in relation to the number of wavelengths used in the calibration and the number of spectra. If a spectrum is listed as a D-outlier, this is the strongest indication that it is a true outlier and has a negative influence on the regression model. Two H-outliers were found in the calibration set based on H>3. They were flagged as outliers simply because they lay at the extremes of the range. In this case, they should not be deleted as they contributed positively to the calibration performance. To make the calibration robust, we left all samples including outliers in the calibration set. To evaluate calibration models, we compared the RMSECV for different sample pretreatments.

The RMSECV is the calibration error and is expressed as

$$RMSECV = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n}} \quad , \qquad (1)$$

where $\hat{y}_i$ are the predicted values out of cross-validation, y_i are the corresponding reference values and n is the total number of samples.

The number of components generating the lowest RMSECV was used in all calibrations to avoid under- and overfitting. The prediction errors (RMSEP) for different treatments were compared with each other in order to evaluate predictions. The RMSECV and the RMSEP are both statistical estimates and are expressed in the same units, being the same measure of the error indicating the type of calibration method used. The coefficients of determination for calibration (R^2) and prediction (r^2), the ratio of performance deviation (RPD) and the ratio of performance to IQ (RPIQ=IQ/RMSECV, with IQ = first quartile subtracted from third quartile of soil sample population) were calculated. The RPIQ was suggested by (Bellon-Maurel et al., 2010) to account for the skewed distribution of soil properties. The 1000 simulations for PLS analysis were performed with orthogonal scores that are implemented in R for Experiment 1 (number of replicates simulation).

3.2 Field-based NIR measurements

3.2.1 Study sites

The study was conducted at three agricultural fields located in Lower Saxony: 'Hinter der Bahn Acker' (HdBA), 'FAL Nord-Ost' (FALNO) and 'Espenberg' (Figure 5). The fields differed partly in soil type (Table 2), and the SOC and total N contents were low. The overall SOC content for the shank reference samples varied between 0.96 and 2.21% and the N-

content between 0.08 and 0.20% (Table 3). The reference samples for probe investigations comprised an overall SOC content from 0.14 to 1.62 % and an N content from 0.02 to 0.16% (Table 3).

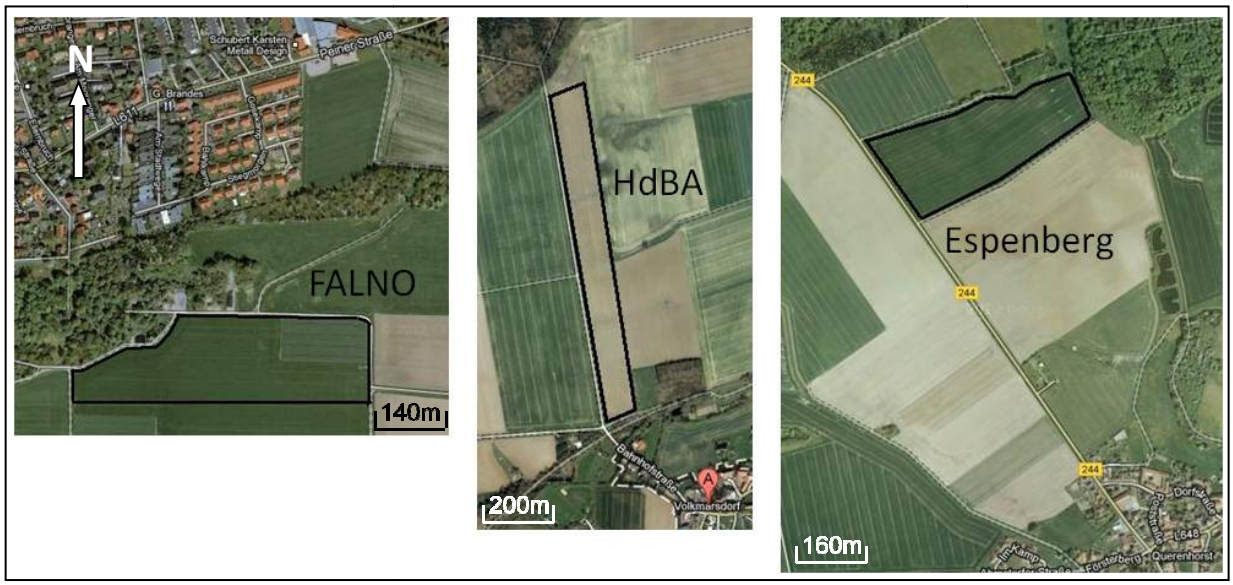

Figure 5: Locations of the investigated fields: FALNO lies 100 m south of Völkenrode at the northwest city boundary of Braunschweig, HdBA 100 m northwest of Volkmarsdorf and Espenberg 900 m north of Querenhorst. HdBA and Espenberg are both ~14 km away from Wolfsburg. The black lines show the contour of the fields.

Table 2: Field information for the FALNO, HdBA and Espenberg fields.

Field	Coordinates	Field size [ha]	Parent rock	Soil
FALNO	10°45´ E, 52°30´ N	9.2	Sandy loess over subglacial till	Stagnic Luvisol (clay = 8%, silt = 28%, sand = 64% →sandy loam)
HdBA	10° 89´ E, 52°36´ N	8.4	Silt over subglacial till	Stagnosol (clay = 20%, silt = 25%, sand = 55% → sandy clay loam)
Espenberg	10°95´ E, 52°34´N	11.6	Silt over subglacial till	Anthric Stagnosol (clay = 15%, silt = 30%, sand = 55% → sandy loam, in parts very calcareous)

Table 3: SOC and total N concentrations for different soil depths of FALNO, HdBA and Espenberg and the results of reference sampling derived from soil cores (for probe measurements) and topsoil composite samples (for shank measurements). The first column shows the mean (X) and the second column the standard error (SE).

Soil sampling for	Depth [cm]	SOC [%]			N [%]		
		FALNO X SE	HdBA X SE	Espenberg X SE	FALNO X SE	HdBA X SE	Espenberg X SE
shank	~ 7	1.34 0.045	0.81 0.03	1.48 0.104	0.11 0.003	0.08 0.003	0.14 0.01
probe	0-10	1.39 0.06	0.76 0.03	1.46 0.19	0.12 0.004	0.07 0.003	0.14 0.02
	10-20	1.42 0.14	0.87 0.03	1.26 0.19	0.12 0.004	0.08 0002	0.12 0.02
	20-30	1.16 0.09	0.72 0.03	0.97 0.13	0.09 0.004	0.07 0.003	0.08 0.01
	30-40	0.46 0.03	0.26 0.02	0.56 0.13	0.05 0.002	0.03 0.002	0.05 0.01
	40-50	0.33 0.03	0.17 0.01	0.46 0.14	0.03 0.003	0.02 0.002	0.04 0.008
	50-60	0.25 0.03	0.15 0.01	0.44 0.16	0.03 0.003	0.02 0.002	0.06 0.015
	60-70	0.19 0.03	0.14 0.01	0.43 0.20	0.02 0.002	0.02 0.001	0.03 0.005

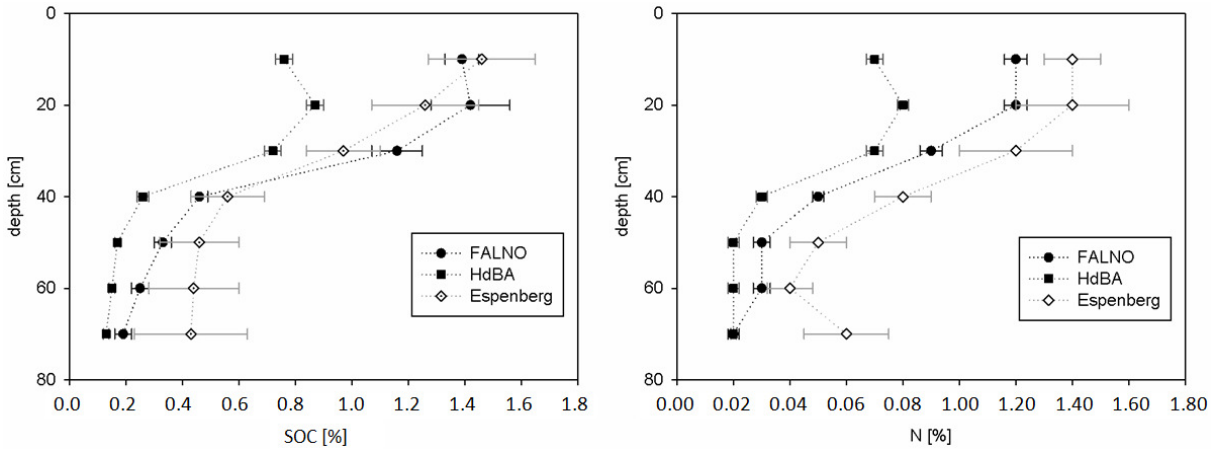

Figure 6: Depth distribution of SOC (left) and N (right) concentrations for a total soil depth of 70 cm for the fields FALNO, HdBA and Espenberg. SOC and N data originated from reference analysis and were averaged per 10 cm depth segments.

The SOC and N content decreased with depth, for the most part (Figure 6). The 10-20 cm depth section held the same or larger average SOC and N concentrations as or than those in the 0-10 cm depth section in four of six cases. This can be mainly attributed to the farmers´ ploughing and the circulating of the field soil.

3.2.2 Mobile field spectrometer

A commercially available system for measuring soil VIS-NIR in the field developed by VERIS Technologies, Inc., Kansas, USA, was used. The complete system encompassed two modules: a) an on-line shank for collecting horizontal VIS-NIR measurements at a discrete depth as it crosses a field (Figure 7) and b) a probe for collecting vertical VIS-NIR measurements of the soil profile to a depth of one metre (Figure 8). During measurement, the optical unit of the shank, which was located at the bottom of the shank, was pulled through the soil by a tractor at an approximately depth of 7 cm. The optical unit had a parallel linkage design, so it followed ground contours precisely, and a toggle-trip design offered protection against rocks and other field barriers. The design provided a dust-free optical path and was self-cleaning. The shank module was additionally equipped with six coulter electrodes that measured soil EC at 0–30 cm and 0–90 cm arrays. The probe collected optical measurements through the sapphire window on the side of the probe as it moved through the soil profile. At the bottom of the probe was a cone-tip with soil EC contacts for collecting dipole EC data.
Two portable cases could be adapted to both measuring systems: a spectrometer case for VIS-NIR measurements and an auxiliary case for EC and GPS measurements. Optical measurements were carried out through a sapphire window fixed on the bottom of the shank and probe, respectively. A tungsten halogen bulb illuminated the soil, and an optic directed the reflected light into a fibre optic cable for transmission to the spectrometer.

Figure 7: Overview of the shank (left): The shank was directly mounted onto a drawbar on the tractor and pulled through the soil. Detail of the shank (right): A chisel opened a trench through which the optical unit got smoothened. Closing discs closed the trench for a nearly planar surface.

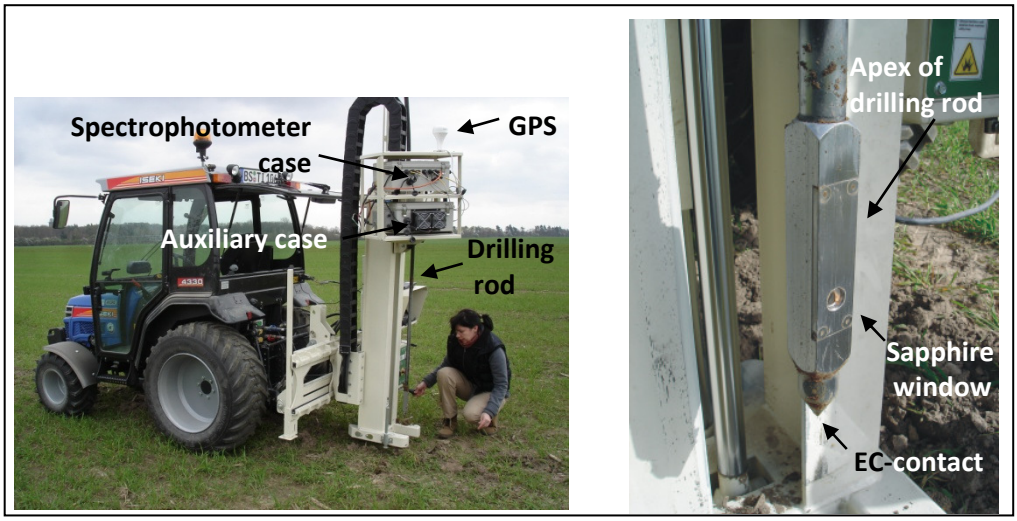

Figure 8: Overview of the probe (left): The probe was directly connected with a tractor and has a white foot as a robust column that connects the hydraulic arrangement with the NIR drilling rod. Detail of the probe (right): It can be lowered just below the soil's surface to a depth of one metre.

The spectrometer was calibrated with four reference grey scale standards before each measurement. The standards had a nominal reflectance of 2%, 10%, 50% and 99%. Additionally, internal shutters of the spectrometer automatically actuated every 15 minutes to collect dark (shutters were closed) and reference spectra (shutters were open – grey scale standards were used). A detailed description of the shank was given by Christy (2008).

The grating spectrometer measured soil reflectance from 350 to 2200 nm with a 6 nm resolution. The reflectance region covered both the visible light and the NIR. One single soil spectrum represented about 3 m of travel. Additionally, GPS data were collected with each measurement; the GPS device was fixed at the level of the sapphire window and was connected with sensor and laptop. After measurement, the data were immediately transferred through a universal serial bus connection to a laptop for storage.

3.2.2.1 Shank: horizontal data acquisition and reference sampling

All shank measurements were performed by driving parallel lines with a distance of 8 m for horizontal data acquisition. The soil sampling technique was i) random distributed or ii) representative orientated (cluster) (Table 5). The latter technique was offered by the VERIS spectrometer software.

Table 4: First results of measurements via shank and probe: number of collected field spectra per field, number of reference samples and sampling technique, driven speed by shank and maximum depth reached by probe.

Field	Shank				Probe			
	Field spectra	Speed [km h^{-1}]	Number reference samples	Sampling technique	Field spectra	Max depth [cm]	Number reference samples	Sampling technique
FALNO	5015	~ 12	36	cluster	3389	70	12	grid
HdBA	6509	~ 11	18	cluster	2038	70	12	random
Espenberg	2305	~ 20	12	random	1669	70	12	random

Here, the entire spectral data input from a field was compressed and then clustered into the desired number of samples. For clustering, a fuzzy-c-means algorithm was used to assign spectra to the cluster with which they had the greatest membership, calculated out of a Mahalanobis distance equation (Christy, 2008). The most representative sample of one cluster was selected by having the greatest number of neighbours geographically and having the lowest distance to the centre. The soil samples at calculated field positions were collected as mixed samples in a length of approximately 3 m and then put in a plastic bag and lastly in a cooling chamber (4°C) until further processing.

3.2.2.2 Probe: vertical data acquisition and reference sampling

The hydraulic probe measurements were taken on a regular grid pattern (24 x 24 m). At each grid point, triple measurements (with a distance of ~25 cm) were performed to catch small-scale variability. Specifically, one main drilling point was first determined, and then two secondary drilling points were set: one left and one right of the main drilling point with a distance of ~12.5 cm. The drilling depth reached a maximum of 80 cm, depending on the soil texture and its resistance to the downward-moving probe. A depth of 70 cm was reached on all fields so that the 0–70 cm depth was set as the general depth under investigation. The insertion speed was around 5 cm per second, and approximately 20 spectra per second were acquired from the spectrometer. In order to avoid countless spectra for the later multivariate data analysis and a possible program breakdown, the software offered the possibility to average the available spectra for different depth sections. To suit to the later reference samples (composite samples for 10 cm depth segments), the spectra were averaged for 10 cm depth sections. According to the grid, 112 main drilling points were scanned at the FALNO site, 97 at the HdBA site and 41 points at the Espenberg site. Soil samples for reference were taken with a cylinder auger (87 mm diameter) down to a depth of 80 cm. Soil cores were then cut into 10 cm segments (starting with 0–10 cm depth) and put separately into plastic bags for transport, before they were temporarily placed in a cooling chamber (4°C). Sampling techniques were i) random distributed or ii) grid orientated (24 x 72 m) (Table 4).

3.2.3 Reference analyses for SOC and N

Moist field soil samples that were collected for reference were dried at 40°C until constant weight. Then, coarse roots and visible stones were removed manually from the sample before the soil samples were directly sieved to 2 mm. Out of the dried and sieved samples, one subsample of 100 g was then taken for grinding in a pebble mill (Retsch GmbH, Haan, Germany, MM400) for 1.2 minutes with a frequency of 30 s^{-1}. Total C and total N concentrations were determined via hot combustion (Vario Max, Elementar Analysensysteme GmbH, Hanau, Germany). For SOC determination, the total C content was subtracted by the SIC. SIC content was defined after ignition at 450°C for 16 hours in a muffle kiln.

3.2.4 Spectra pretreatment

Firstly, all field spectra were converted from wave number [cm^{-1}] into wave length [nm]. For the purpose of figure illustration, the reflectance values were additionally converted to absorbance values as the logarithm of the inverted reflectance spectra, as this spectrum transformation is widely used for the presentation of VIS-NIR spectra in the context of organic matter assessment (Cozzolino and Moron, 2006; Bartholomeus et al., 2008). The scanned spectral regime comprised the VIS and NIR regions. Later, the VIS-NIR spectra were additionally cut into the solely NIR region so that the influence of the VIS-NIR and NIR spectra on calibration accuracy could be compared. A solely VIS investigation was not undertaken since the capability and benefit of NIR was yet to be explored.

For shank calibrations, the NIR range (1100 to 2220 nm) was compared with the VIS-NIR range (400–2220 nm). For probe calibrations, the VIS-NIR range alone was taken for all measurements. Here, the NIR range itself yielded in worse calibration results, with R^2 lower than 0.5 for SOC (HdBA: $R^2 = 0.482$, RMSECV = 0.089 and FALNO:$R^2 = 0.472$, RMSECV = 0.272) and even lower than 0.1 (Espenberg: $R^2 = 0.084$, RMSECV = 0.539) in general. Results in brackets were the highest R^2 values out of all the calibration results with corresponding RMSECV-values. Adding the VIS-range from 400 to 1100 nm to the NIR significantly improved the probe calibration results for HdBA and FALNO ($R^2 \geq 0.9$) and moderately for Espenberg ($R^2 \geq 0.4$). Of course, the R^2 value is not a good sole criterion for calibrations, since it depends on the measurement range (Davies and Fearn, 2006). However, as stated by Couteau and Schaller (2003), it gives information as to whether a quantitative calibration is acceptable ($R^2 > 0.8$) or not ($R^2 < 0.8$).

Several spectra manipulations were used to identify the influence on calibration accuracy: Calibrations were carried out with reflectance spectra (refl) and with those converted into first (1^{st} der) and second Savitzky-Golay derivative (2^{nd} der), standard normal variate (SNV), detrend, normalisation (Normal) (Otto, 1997) and smoothing. Generally, spectral derivatives remove baseline effects, described as the constant underlying spectra and other non-systematic effects that influence the global shape and the absolute level of a spectrum (Heise and Winzen, 2006). The first derivative determines the slope of the spectral curve at each point, and the second derivative computes the change in the slope of the curve, removing trends if there are any (Duckworth, 2004). The standard normal variate method removes the major effects of light scattering from the spectra (Duckworth, 2004). Detrend is recommended when background interference is not greatly influenced by chemical properties, since it removes variations useful for modelling. However, it fits a polynomial of a given order and

subtracts this polynomial and is therefore useful for curve offsets. Normalisation is used to identify and remove sources of systematic variation between spectra, such as variation in the instrument detector sensitivity and also variation in certain properties of the soil sample that are not attributed to soil C. Smoothing assumes that variables near to each other in the data matrix are related to each other and therefore contain similar information. This information is averaged to reduce noise without a significant loss of the spectral signals.

3.2.5 Calibration model development

The Calibration Wizard version 1.1 (SensoLogic GmbH, Norderstedt, Germany) was used to build calibration models. The calibration process comprised correlating the SOC and total N content to the calibration samples with their spectral data recorded in the field. Calibrations out of probe measurements via a regular grid comprised a correlation of three spectra per grid point with their reference SOC and N content. Then, PLSR were carried out using a leave-one-out cross-validation. PLSR is basically linear, and it can analyse data with noisy and numerous variables as well as simultaneously model several response variables (Wold et al., 2001). The complexity of PLSR is controlled by the number of PLSR components. Leave-one-out cross-validation (Efron and Tibshirani, 1994) determines the optimum number of components by estimating the final lowest RMSECV (Viscarra Rossel et al., 2006; Brunet et al., 2007). The optimum ranged between seven to 10 components in all SOC calibrations and between five to nine components in all N calibrations.

Spectral outliers were removed, according to Shenk and Westerhaus (1991) and Barthes et al. (2006), when they had a Mahalanobis distance greater than three, since these spectra deviated far from the mean sample spectrum. The maximum number of removed outliers in a calibration was two. The best spectral treatment was the one having the highest R^2 and the lowest RMSECV. This is in accordance with Knadel et al. (2011), who ranked their validation results for best calibration results showing lowest prediction error and highest coefficient of determination. Listed in the tables were consistently the best calibration results. With regard to the reproducibility study, different calibration models were also rated according to the ratio of the standard error of calibration (SEC) to the mean reference value (SEC [%]) as well as the residual prediction deviation (RPD), which is the ratio of standard deviation of the measured SOC values to RMSECV, and the ratio of error range, which is the ratio of the reference data range to RMSECV (RER). The SEC, RER and RPD values were added, since these values include information from the reference data set, an important data basis for investigating the reproducibility of the SOC calibrations. Additionally, an off-line prediction

with Calibration Utilities version 2.0 (SensoLogic GmbH, Norderstedt, Germany) was carried out, and the SOC data of all field spectra per study and sampling campaign could be predicted on the basis of the best calibration model. To distinguish between the SOC data derived through laboratory measurements from reference soil samples and the SOC data predicted from NIR field spectra, the SOC data from the laboratory measurements were additionally named with REF, and predicted SOC data were additionally named with TEST.

3.2.6 Geostatistical models for data interpolation

Management practices in precision agriculture require accurate contour maps of soil properties. It is well known that Kriging produces reliable predictions for mapping. Nevertheless, enough sample data are required in order to build an accurate semivariogram. To do so, we expanded our reference field data with a simulated 4 x 4 m grid-layer. Via Kriging, SOC values were predicted for all intersections. Eight models for data interpolation were used to determine which fit best – a spherical, an exponential, a Gaussian or an M. Stein's parameterisation model – all with and without nugget (all non-linear regressions). The eight models were available through the 'autokrige' package from R Studio (R Core Team, 2013), version 2.15.2, which was used for statistical calculations. Out of all predicted SOC values via Kriging, a median was calculated that represented the spatial weighted field SOC stock. The maximum distance of all semivariograms was set to 300 m, for the 'field-based NIR measurements' as well as the 'field reproducibility study'. Fixed distance intervals of 4 m were selected. The results included data for sill, range and nugget. In general, the sill determines the limit of the semivariogram tending to infinity lag distances, the range is the distance in which the difference of the semivariogram from the sill becomes negligible, and the nugget is the height at which the semivariogram cuts the Y-axis. The determination of the best model was based on that with the lowest Akaike value (AIC).

3.2.7 Calculation of SOC stocks

SOC reservoirs are monitored for climate impact assessment, as they are not permanent and react sensitively to global warming. With the knowledge of SOC content and bulk density (BD), the SOC stock could be calculated (equation 2). Consequently, SOC concentration and SOC stock are not independent variables, but they are correlated with each other. Nevertheless, SOC stocks give a better understanding of the carbon content per a given soil volume since disturbing effects such as soil pores and stones are minimised, which is the reason for a continuous work with SOC stocks was preferred.

SOC stocks were generally calculated for the topsoil, covering a depth of 0–30 cm, as follows:

$$SOC_{stocks} = \sum_{i=1}^{n} BD_{finesoil,i} \times SOC_{conc,i} \times depth\_volume_i . \quad (2)$$

The depths involved (i) were 0–10 cm, 10–20 cm and 20–30 cm, and n was the number of sampled soil depth intervals. BD finesoil,i and SOCconc,i represented the bulk density and SOC concentration of the fine soil, respectively.

3.2.7.1 Determination of bulk density

Bulk density is a physical soil parameter and is used to quantify soil compactness and to calculate SOC stocks. BD varies with soil management as well as with inherent soil properties, and it generally decreases with an increase in organic matter and moisture content. For BD assessment, coarse roots and plant residuals (> 2 mm) were extracted by hand from the fresh samples. The air-dried soil was weighed and then dried in a drying chamber with 40°C for 24 hours until reaching constant weight and was then weighed again and sieved to 2 mm. The sieved samples were weighed again to determine the stone mass. The density of stones (p(stones)) was approximated by 2.63 g cm^{-3} according to Kretzschmar (1986). This is the density of quartz, which was assumed to be the principal component of the stones in the sand dominated soils. Additionally, the bulk density was corrected for free water in the air-dried samples.

The BD of the soil was calculated as follows:

$$BD = \frac{mass(sample) - mass(stones)}{volume(sample) - \left(\frac{mass(stones)}{p(stones)} \right)} \quad (3)$$

BD could be directly calculated out of the reference soil samples taken via Eijkelkamp auger, since the size and volume of the auger was known as well as the sampled depth (for calculation of 'volume(sample)'). These BD data were called BD_{MEAS}.

The BD data of the composite soil samples could not be calculated via equation 2. Here, the soil was sampled by a scoop without any DIN standard for soil volume. However, it is common to find databases or datasets worldwide that lack direct BD measurements for all or some records due to time-consuming and expensive investigations (Sequeira et al., 2014). This need for BD data has led to the development of a variety of pedotransfer functions

(PTFs) that predict BD using information from more easily obtainable and available data, including physical and chemical soil properties such as soil texture, C and pH (Adams, 1973; Rawls, 1983; De Vos et al., 2005; Heuscher et al., 2005). Since we measured the SOC content of all soil samples, we searched for PTFs that included solely this variable. Six PTFs were checked for best suitability (Table 5). Out of the six PTFs, the ones from Callesen et al. (2003) were most appropriate: They included a large number of tested soil samples, R^2 over 0.5 were satisfactory (Chang et al., 2001) and the tested soil characteristics directly met the texture of our fields, which were sandy (FALNO, 64% sand proportion) or loamy sand / sandy clay loam soils (HdBA and Espenberg, both 55% sand proportion). Therefore, regressions five and six from Callesen et al. (2003) were used for BD calculation: five for FALNO and six for HdBA and Espenberg. The BD data derived from these PTFs were called BD_{CALL}.

Table 5: Pedotransfer functions and their statistical background, comprising the number of tested soil samples, the number of soil types used and the R^2. Regressions were selected out of following documents: 1) Manrique and Jones (1991), 2) Alexander (1980), 3) Federer (1983), 4) Huntington et al. (1989) and 5+6) Callesen et al. (2003). NM = not mentioned in the paper. OM = organic matter.

Regression no.	1	2	3	4	5	6
Function	BD= 1.66 - 0.318 √(SOC)	BD= 1.66 - 0.308 √(SOC)	ln(BD) = -2.31 - 1.079 ln(OM) - 0.113 $[\ln(OM)]^2$	ln(BD) = -2.39 - 1.316 ln(OM) - 0.167 $[\ln(OM)]^2$	BD = 1.59 - 0.105 √(SOC)	BD = 1.83 - 0.131 √(SOC)
No. of tested soil samples	19651	721	480	60	844	844
Soil types or soil characteristic	9 different soils	7 different soils	4 forest soils	course loamy soils	sandy soils	loamy sand
R^2	0.41	0.46	NM	0.75	0.66	0.58

Even though the BD data of the soil cores could be determined by equation 2, an additional BD assessment with Callesen et al.'s PTFs was carried out in order to compare the results of BD values and SOC stocks. Supplementary field-specific regressions for FALNO, HdBA and Espenberg were developed by plotting SOC concentrations (SOC data from direct chemical analysis) with corresponding BD data (BD_{MEAS} where available, otherwise BD_{CALL}). A linear regression equation of the point cloud diagram was used as a field-specific BD equation (e.g. 0–10 cm depth: FALNO, BD = -0.108 x (SOC) + 1.673, HdBA, BD = -0.29 x (SOC) + 1.919

and Espenberg, BD = -0.126 x (SOC) + 1.851). Field-specific BD data could be obtained after having inserted the SOC data into the equation. For each depth segment of 10 cm, a BD regression was computed. The calculated BD data were called BD_{FS}. Summing up, three types of BD values (calculations) were used for comparison.

3.2.7.2 Calculation of SOC stock models

Since SOC stocks are an important issue in the context of climate change, numerous publications deal with this topic. However, the magnitude of SOC stocks depends on the variables involved, such as SOC concentration, bulk density, sampling depth and rock fragment content. Considering BD, the magnitude of SOC stocks is also dependent on how the BD value was determined. This is not consistent throughout the literature. To visualise the differences in SOC stock data arising out of varying calculations, fourteen different SOC stock models were computed. They differentiated in calculations in terms of on-line NIR mapping (six models, Table 6) and direct chemical analysis in the laboratory (eight models, Table 7), as well as the kind of soil samples used – composite or soil core samples. Along with three types of BD assessment (mentioned in 3.2.7.1), two different calculations for the depths were used for SOC stock estimation:

$$SOCstocks(0-30) = BD_{finesoil(0-10,10-20,20-30)} \times SOC_{conc(0-10,10-20,20-30)} \times depth\_volume_{(30)} \quad (4a)$$

$$SOCstocks(0-30) = [BD_{finesoil(0-10)} \times SOC_{conc(0-10)} \times depth\_volume_{(10)}] \times 3 \quad (4b)$$

In equation (4a), the soil data of three depth sections were used. For equation 4b data solely from the upper depth section (0–10 cm) were used and then extrapolated to 0–30 cm. This was applicable for soil core samples. Composite samples that were collected at a depth around 7 cm were handled as follows: Since the estimated SOC content at a depth of 7 cm was assumed to be representative of the average SOC content at 0–10 cm, the soil depth was set to 0–10 cm for all further calculations.

Table 6: Indication of 6 SOC stock models that were based on on-line NIR measurements. Two types of BD assessments (BD_{FS} and BD_{CALL}) combined with two depth calculations were used for SOC stock assessments from composite and soil core samples. Data for depth calculations were from 0–10 cm (equation 4b) and 0–30 cm depth (equation 4a).

Sampling operation	composite sample		soil core			
SOC stock models	A	B	C	D	E	F
Bulk density	BD_{FS}	BD_{CALL}	BD_{FS}	BD_{CALL}	BD_{FS}	BD_{CALL}
SOC concentration	NIR					
sampling depth [cm]	0-10	0-10	0-30	0-30	0-10	0-10

Table 7: Indication of 8 SOC stock models that were based on direct chemical analysis of SOC concentration from soil samples (no NIR mapping). Three types of bulk density assessments (BD_{FS}, BD_{CALL} and BD_{MEAS}) combined with two depth calculations were used for SOC stock assessments, from composite and soil core samples. Data for depth calculations were from 0–10 cm (equation 4b) and 0–30 cm depth (equation 4a).

Sampling operation	composite sample		soil core					
SOC stock models	G	H	I	J	K	L	M	N
Bulk density	BD_{FS}	BD_{CAL}	BD_{FS}	BD_{CAL}	BD_{MEA}	BD_{FS}	BD_{CAL}	BD_{MEA}
SOC concentration	Direct chemical analysis							
sampling depth [cm]	0-10	0-10	0-30	0-30	0-30	0-10	0-10	0-10

3.2.8 Error analysis of SOC stocks

The Kyoto Protocol and the European Common Agricultural Policy rely on C stock assessment as part of the verification of changes in soil organic matter. Large uncertainties of SOC stock estimation exist, which may impair the detection of temporal SOC stock changes. These uncertainties are rarely quantified even though they are critical in determining the significance of the results. Five error values were examined with reference to SOC stock model assessment.

First, the standard error SE_{REF} (REF = direct chemical analysis) was calculated out of all SOC stocks at reference sampling positions (n_ref) together with their standard derivation (SD):

$$SE_{REF} = \frac{SD}{\sqrt{n\_ref}} \tag{5}$$

According to Lischer (1993) and (Vandervaere et al., 1994), the measurement error (SE_{MEAS}) was determined to see its general proportion on the standard error. For calculation, the error from SOC-analysis (Vario Max, Elementar Analysensysteme GmbH, Hanau, Germany) was

taken into account as well as the BD error and the covariance between SOC and BD (equation 6). The standard error of BD (σBD) was 0.047 g cm^{-3} (pers. comm. Axel Don). Callesen's BD error was 0.18 g cm^{-3} for sandy soils (to be applied to FALNO) and 0.15 g cm^{-3} for loamy sand and sandy loam soils (to be applied to HdBA and Espenberg):

$$SE_{MEAS} = SOCstock \times \sqrt{\left(\frac{\sigma SOC^2}{SOC^2}\right) + \left(\frac{\sigma BD^2}{BD^2}\right) + \left(\frac{\mathrm{cov}(SOC,BD)}{SOC \times BD}\right)^2} \quad (6)$$

Both SE_{REF} and SE_{MEAS} were used for SOC stock models A–F, which were based on direct chemical analysis (laboratory) without on-line NIR measurements.

For SOC stocks based on NIR field data (SOC stocks G-N), three types of errors were analysed. The standard error of SOC stocks SE_{PRED} (PRED = predicted SOC-values out of calibration) at mapped field positions (n_map) together with the standard derivative was calculated as follows:

$$SE_{PRED} = \frac{SD}{\sqrt{n\_map}} \quad (7)$$

Next, the standard error SE_{RMSEP} was computed using the error out of calibration. Hereby, the assumption was made that the calibration error (RMSECV) was not significantly different from the prediction error (RMSEP) so that RMSECV was equated with RMSEP and was used for calculation. In this connection, the number of mapped field positions was corrected for its range. The range was determined via ordinary Kriging of SOC stock data and their field coordinates (see 3.2.6). The range indicates the distance at which field data are spatially dependent on or independent of each other. According to Carmelino Hurtado et al. (2009), to guarantee spatially independent data, the actual number of mapped field positions was corrected for error analysis as follows:

$$n\_map(range\_corrected) = \frac{n\_map}{(range \times distance\_of\_spectral\_scanning)} \quad (8)$$

$$SE_{RMSEP} = \frac{RMSEP}{\sqrt{n\_map(range\_corrected)}} \quad (9)$$

In order to have a comparative error to the SE_{REF}, the standard error of a Monte Carlo simulation (SE_{MC}) was developed. Both errors took the number of reference sampling points into account. For SE_{MC}, each studied field was divided into plots of equal size. The number of

plots was equivalent to half of the number of reference sampling points. Each sampling point was described by the value of its carbon content. Via a Monte Carlo simulation, two SOC values were selected from each plot. The values were put back into the plot until all possible combinations of two SOC values were performed. Then, the mean SOC value was calculated for each plot, followed by computing the median out of all means. Lastly, the standard error (SE_{MC}) was calculated.

3.2.9 Time aspect of SOC stock estimations

There is an increasing demand for establishing new technologies with respect to a time- and cost-effective way of handling instruments and statistical procedures. Therefore, the time factor 'how long does it take to do what' was set as an important criterion for further SOC stock evaluations. Here, Table 8 gives an overview of the relevant workspaces, necessary tasks and their amount of time, the operators and the salaries. All notes were stated for an average field with a size of 10 ha. For all tasks, optimum measurement conditions were assumed, such as no technical problems and good weather conditions. Times of travel were not considered. For the employers' gross wages to be paid, a standard salary for research assistants of 4800 € per month, a salary for technical assistants of 3200 € per month and an hourly wage of 13 € for student assistants were applied.

For further calculations, the gross wage in Table 8 also included overhead costs of 120%, such as costs for offices, rooms, facilities, equipment and services. Some aspects were not listed in the table, such as performing calibration and validation, BD analysis and data interpolation, since they were regarded as minor time factors. However, the training of the personnel consumed some time, and the time amount can differ from personnel to task. Therefore, it was not integrated into our analysis, yet the time should be considered when a new project is started. Overall, the calculation did not represent a detailed business cost analysis since the expected useful lives and running costs of the equipment (e.g. the servicing, calibration) were not known. However, these factors should not be considered in the context of this study. For the assessment of all SOC stocks, the total amount of working time (laboratory, field and office work) was estimated. The calculated working time was plotted against the value of the standard error of SOC stocks to determine whether a correlation between time and standard error exists.

Table 8: Tasks for SOC stock estimation of a 10 ha field in relation to workspace, time and type of operator. Operators are student assistants (STA), technical assistants (TA) and research assistants (RA) listed with their corresponding employers' gross wages to be paid.

Workspace	Tasks	Tasks valid for	Working time[h]	Operator	Gross wage
Field	(1) NIR data acquisition: shank	10 ha field → ~ 11 km driving line	3	TA	135.30 €
	(2) NIR data acquisition: probe	10 ha field (24 m x 24 m grid) → 168 point measurements	4	TA	180.40 €
	Soil sampling for reference analysis: topsoil (1)	12 soil samples	0.75	STA	21.45 €
	Soil sampling for reference analysis: soil profile (2)	12 soil cores, (with 7 samples each)	4	STA	114.40 €
Laboratory	Sieving, grinding, SOC-/ N-analysis (1)+(2)	12 soil samples , 0.5h / soil sample	6	TA	270.60 €
	Drying of probe samples for SOC-/ N-analysis (2)	12 soil samples	0.2	STA	5.72 €
	Weighing before and after drying of samples (2) for BD estimation	12 soil samples	0.5	STA	14.30 €
	Drying of shank samples for SOC-/ N-analysis (1)	12 soil samples	0.2	STA	5.72 €
Office	Generation of a field specific BD formula, quality assurance	all above mentioned operations	3	RA	203.08 €

3.3 Field reproducibility study

3.4.1 Study sites

The study was performed at two agricultural fields located in Lower Saxony: 'FAL Number 250' (FAL250) and 'Hinter der Bahn Acker' (HdBA) (Figure 9). HdBA was discussed in section 3.2.1 dealing with the comparison between horizontal and vertical on-line mapping. The present fields differed in soil type and texture (Table 9). According to the FAO soil classification, the texture of FAL250 was classified as sandy loam and that of HdBA as sandy clay loam.

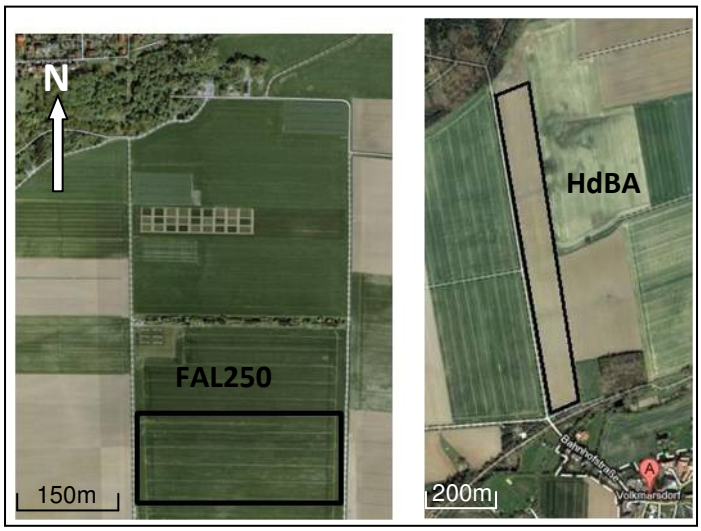

Figure 9: Locations of the investigated fields: FAL250 can be found 600 m south of Völkenrode (at the northwest city boundary of Braunschweig), and HdBA is 14 km southwest of Wolfsburg. Black lines show the borders of the fields.

Table 9: Field information for FAL250 and HdBA.

Field	Coordinates	Field size [ha]	Soil type	Parent rock	Soil texture [%]		
					Sand	Silt	Clay
FAL250	10°45´ E, 52°29´ N	4.25	Stagnic Luvisol	Sandy loess over subglacial till	64	27	9
HdBA	10° 89´ E, 52°36´ N	8.40	Stagnosol	Silt over subglacial till	55	25	20

Table 10: Reference data sets of HdBA and FAL250 according to repeated soil sampling within two measurement campaigns on both fields: minimum, maximum and mean SOC concentration of soil samples, their standard deviation SD, standard error SE and skewness. Single SOC concentrations, needed for calculation of the mean, were measured in a laboratory and used for SOC calibration.

Fields	measurement campaign -1-						measurement campaign -2-					
	SOC g kg^{-1}						SOC g kg^{-1}					
	Min	Max	Mean	SD	SE	Skew-ness	Min	Max	Mean	SD	SE	Skew-ness
FAL250	11.5	19.5	14.7	1.8	0.5	0.99	11.3	18.1	15	1.8	0.4	0.79
HdBA	5.9	11.1	8.1	1.3	0.3	0.31	7.9	9.9	9.3	0.6	0.2	-1.39

The study encompassed repeated measurement campaigns that included on-line NIR measurements, soil sampling at defined field positions and laboratory analysis of SOC content. The SOC concentrations of both fields were low (Table 10). The distribution of SOC

concentrations was skewed to the left in three of four cases and once, for the second measurement of HdBA, it was skewed to the right.

3.4.2.1 Data acquisition and soil sampling

Horizontal field NIR measurements were carried out using the VERIS shank. The type of measurement and soil depth studied were the same as discussed in section 3.2.2. Each field measurement was conducted continuously by driving in semicircles at the field's end for an s-shaped lines optic. The average speed driven and number of collected field spectra as well as the dates of repeated measurement campaigns can be seen in Table 11. The measurements were carried out directly after harvesting. Before the start of the second measurement campaign of HdBA (date = 09.10.12), the soil was ploughed by the farmer.

The reference soil samples needed for the determination of the fields' average SOC content and for SOC calibration were collected at field positions that should deliver representative information of SOC concentrations from FAL250 and HdBA. The sampling strategies for the reference samples were cluster analysis and a nested design (Table 12).

Table 11: Number of field spectra recorded via the NIR shank within the first and second measurement campaigns on the fields FAL250 and HdBA. The dates of NIR measurements as well as of reference soil sampling are added together with the number of cleaned field spectra and averaged speeds driven. Cleaned field spectra are filtered field spectra without outliers. The VERIS spectrometer software removed the outliers automatically after a principal component compression of all field spectra per measurement campaign.

Fields	measurement campaign -1-				measurement campaign -2-			
	Number of		Driven speed [km h^{-1}]	Dates of NIR measure-ments/ soil sampling	Number of		Driven speed [km h^{-1}]	Dates of NIR measure-ments/ soil sampling
	Field spectra	Cleaned field spectra			Field spectra	Cleaned field spectra		
FAL250	3086	1227	~ 13	27.03.12	2201	2136	~ 14	10.04.12
HdBA	6509	6288	~ 11	03.09.10	1870	1582	~ 15	09.10.12

The cluster analysis is also called 'analysis for the most representative location' since it incorporated the spectral information from each field NIR measurement. The clustering function was directly implemented in the VERIS spectrometer software, and it compressed the whole spectral data input from the measurements of HdBA and clustered it. The sampling point with the greatest number of geographic neighbours was then selected out of each cluster (see section 3.2.2.1 and 3.2.5).

Table 12: Number of reference soil samples collected from the fields FAL250 and HdBA within the first and second measurement campaigns. Soil samples were composite samples collected via two strategies: nested and cluster designs.

Fields	measurement campaign -1-		measurement campaign -2-	
	Number of reference samples	Sampling strategy	Number of reference samples	Sampling strategy
FAL250	19	nested	19	nested
HdBA	18	cluster	18	cluster

The nested sampling design, developed by Youden and Mehlich (1938), was applied to establish an adequate basis for Kriging – a geostatistical method that can describe and predict the spatial variation of SOC. The design was modified as follows: Two rectangles (thick dashed lines) were imaginarily placed on the field FAL250 (Figure 10). Their distance to the nearest field boundary was 5 m, and for each rectangle, 9 sampling points were defined.

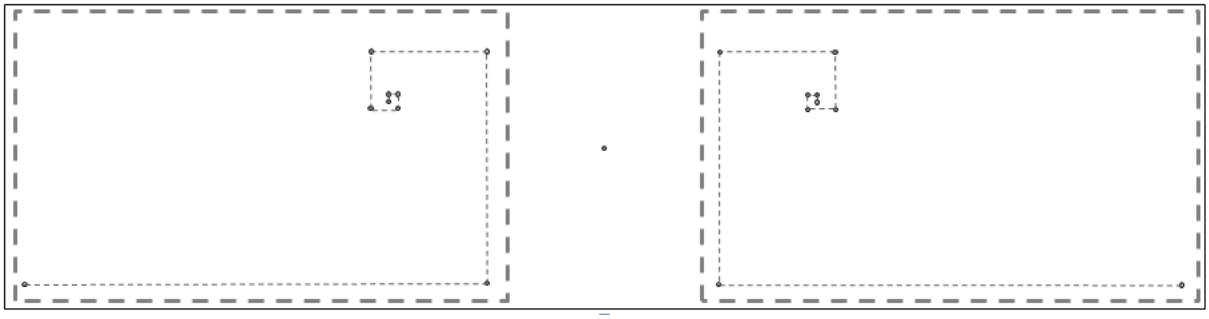

Figure 10: Sketch of 19 reference sampling positions on the field FAL250 (black solid contour line), following a nested design.

Starting with the first point, the smallest distance to the second point was set at 8 m – the driven line distance of the NIR measurement. The sampling positions followed a previously defined field quadrant with directions X and Y (thin dashed lines). The distances between sampling points were 8 m, 16 m, 32 m, 64 m, 128 m, 256 m, 512 m and 1024 m. The last point was set at the field's corner with a distance of 8 m to the field boundaries. The first rectangle was mirrored at a line with north-south orientation, which halves the field into a second rectangle for a double dataset. One additional sampling point was set in between the two rectangles. Reference soil samples were directly collected from the shank driving path. Since one soil spectrum represented about 3 m of travel (see section 3.2.2), calibration soil samples were collected as mixed samples in a length of approximately 3 m. Soil sampling and NIR measurements were carried out twice on each field and were later referred to as the first

(-1-) or second (-2-) soil sampling or NIR measurement campaign, respectively. Additionally for FAL250, the SOC data from the reference soil samples (first measurement campaign) were combined with the spectra from a second NIR measurement for SOC calibration, assuming no change of SOC within two weeks. The measurement was referred to as measurement type two with repeatedly used soil SOC data (-2- ^{rep}). This measurement type was used as an indicator to emphasise the influence of repeatedly scanned field spectra on calibration accuracy and predicted SOC concentrations and SOC stocks when compared to measurement -1-. All collected soil samples were put in a plastic bag and then put in a cooling chamber (4°C) until further processing.

3.4.2.2 Spectral data and calibration

According to section 4.2.1 and to successful calibration results with $R^2 \geq 0.9$ for SOC, spectra were cut into the NIR wavelength range from 1100 to 2220 nm. The use of the complete VIS-NIR range was put aside due to the worse SOC calibrations results with $R^2 \leq 0.82$. Spectra pretreatment and calibration model development were performed as described in sections 3.2.4 and 3.2.5. For comparison purposes, one H-outlier was removed from the reference calibration set of each measurement campaign on HdBA, since each H-outlier had a great influence on each calibration model with a large distance from the soil spectrum to the regression line (measurement one: with outlier→ R^2=0.91 [see Table 19], without outlier→ R^2=0.99; measurement two: with outlier→ R^2=0.93, without outlier→ R^2=0.99).

3.4.2.3 SOC stock calculation

Increased interest in global estimates of SOC stocks is due to the potential effects of climate change on agricultural productivity and the carbon stored in soils. In order to quantify the SOC stocks of FAL250 and HdBA, we converted carbon into carbon stocks. SOC stocks were calculated for the topsoil, covering a depth of 0–30 cm, as was done in section 3.2.7.2 via formula 4b with an extrapolation from 0–10 cm to a 0–30 cm depth. SOC data were available from laboratory analyses. Since composite soil samples were collected in the field, BD data for the samples were not available. That is why the BD of the soil was estimated by pedotransfer functions. In a first step, a regression from Callesen et al. (2003) was used. It was suitable for our fields (HdBA and FAL250) with sandy soil: $BD = 1.59 - .105 \sqrt{(SOC_{conc})}$. In a second step, field-specific regressions were used. Their generation and formulae are listed and described in section 3.2.7.1. The field-specific BD regression for HdBA was adopted from this section: $BD = -0.29 \times (SOC_{conc}) + 1.919$. For FAL250, the BD regression from

FALNO was applied: Both fields were geographically very close to each other, with a minimal distance of about 400 m. That is why both fields were assumed to differ marginally in soil characteristics. The regression used was: BD = -0.108 x (SOC_{conc}) + 1.673. Estimated SOC stock data were described via the mean value, the standard error and the variation coefficient.

3.4.2.4 Calculation of minimum detectable difference (MDD)

SOC stocks in soils are not permanent, since the soil is influenced by global warming as well as by mechanical soil disturbances, changes in plant and vegetation growth and many other factors. Before changes of SOC content can be clearly detected, it is essential to determine the predictive accuracy of on-line NIRS investigations in general. This was done via the use of the MDD, which displays the results of SOC changes that are theoretically detectable. Results from repeated measurement campaigns can then be compared for their reproducibility.

Following the recommendations of Krebs (1999), the standard error (SE), the type I error α with $z_{\alpha}^{(2)}$, which is the two-sided critical value of the normal distribution at a given significant level, and $z_{\beta}^{(1)}$, the one-sided quartile of the normal distribution corresponding to a probability of type II error β, were needed for MDD assessment via the usual approximation formula:

$$MDD \geq \sqrt{2 \times SE\left(z_{\alpha}^{(2)} + z_{\beta}^{(1)}\right)} \qquad (10)$$

Additionally, the SE of the mean of all SOC concentrations or the SE of the mean of all SOC stocks per field and measurement campaign was inserted into equation 10, dependant on whether the MDD of SOC concentrations or the MDD of SOC stocks was to be calculated.

Type I error α generally reflects the level of significance and the interval in which the expected value will be with a probability of 95%. By convention, α is often set to 0.05 (α=0.05). Considering the number of reference samples of 17 (HdBA, after outlier removal, see Table 25) and 19 (FAL250) as well as the number of field spectra of >1000 (for HdBA and also FAL250), the type I α error for reference samples was $z_{\alpha}^{(2)}$=2.10, and for field spectra it was $z_{\alpha}^{(2)}$=1.96.

Type II error β of 0.2 (β=0.2) corresponds to a statistical power of 0.8. This power is generally regarded as sufficient. In regard to the number of reference samples and the number of field spectra, the type II β error for reference samples was $z_{\beta}^{(1)}$=0.86, and for field spectra it was $z_{\beta}^{(1)}$=0.84. Both error types were obtained from the Students t-table.

Equation 10 referred to two unpaired samples sets. Repeated measurements of HdBA and FAL250 comprised repeated soil sampling and repeated NIR measurements. Even though the aim was to resample and re-measure at similar field positions, this cannot be ensured with accuracy to 1 mm. That is why unpaired sample sets were regarded as applicable. Equation 11 referred to two paired sample sets. For FAL250, SOC concentrations from soil samples that were collected within the first sampling campaign were used for MDD calculation, as they were SOC data from a second measurement campaign, assuming a negligible or no change in SOC within a time interval of two weeks.

$$MDD \geq \sqrt{1 \times SE\left(z_{\alpha}^{(2)} + z_{\beta}^{(1)}\right)} \tag{11}$$

For each field and measurement campaign, the MDD was calculated six times, twice for organic carbon concentration (NIR mapping vs. conventional sampling) and four times for SOC stocks (NIR mapping vs. conventional sampling, Callesen-BD vs. field-specific-BD). In addition to the mentioned absolute MDD, the relative MDD was calculated by dividing the absolute MDD by the mean value of the SOC concentration or SOC stock dataset per field and measurement campaign. Next, in order to avoid field points with equal or nearly equal Kriging weights within short distances, a range-correction was carried out that reduced the number of field spectra for MDD assessment. Only uncorrelated field points were used for calculation. In a first step, the range from Kriging statistics was needed to generate the number of field spectra within the range:

$$\frac{range[m]}{3[m]} = number\ of\ spectra\ within\ range. \tag{12}$$

For this purpose, the range was divided by 3 m – the measuring length that is needed to obtain one (average) field spectrum. Next, all field spectra per field and measurement campaign (cleaned field spectra) were divided into the number of field spectra within their range. The results were the number of uncorrelated field spectra needed for calculation of the SE of SOC concentrations and SOC stocks for MDD estimation.

In a next step, the MDDs of SOC concentrations were assessed in relation to the number of reference soil samples. In order to do so, a Monte Carlo simulation was performed with R Studio (R Core Team, 2013) version 2.15.2 and package ‘mc2d’: The simulation varied the number of existing reference samples for each field: between 1 and 17 (after outlier removal, see Table 25) for HdBA and between 1 and 19 for FAL250. Corresponding to the number of samples, SOC values were selected from the whole sample set, and a median was computed.

Then, different combinations of samples were selected until all possible combinations were performed. The MDD was determined separately for the fields FAL250 and HdBA as well as for each measurement campaign.

In a last step, the theoretical time difference of the two studies was calculated, which detects the smallest difference of a SOC stock with statistically significance. Ciais et al. (2010) inferred a loss of SOC stock in cropland soils of 0.17 Mg C ha^{-1} yr^{-1} as the European average. Assuming a constant SOC stock change over the years, their modelled SOC stock loss and the MDDs calculated in this study served as a basis for the computation of the time difference or the number of years, respectively.

4 Results

4.1 Laboratory-based NIR measurements

4.1.1 Effect of repeated measures on NIR calibrations

With respect to RMSECV and R^2, calibration results for SOC and also total N were similar and no statistical differences could be found for predicted values out of cross validation ($p<0.05$) for sieved air-dried samples when more than 1 sample was scanned. However, there was a slight trend of decreasing calibration error with increasing number of replicates (Table 13).

Table 13: Calibration results of SOC and N: one to four measurements per sample of sieved soil were performed; results were compared with twice measured ground samples.

Sample pretreatment	Number of replicats	SOC			N		
		RMSECV	R^2	RPIQ	RMSECV	R^2	RPIQ
sieved, air-dried	1	0.583	0.601	2.75	0.045	0.673	1.99
sieved, air-dried	2	0.583	0.601	2.75	0.045	0.675	1.99
sieved, air-dried	3	0.582	0.603	2.77	0.045	0.682	1.99
sieved, air-dried	4	0.581	0.603	2.77	0.044	0.686	2.03
ground, air-dried	1	0.453	0.812	4.58	0.034	0.864	8.94
ground, air-dried	2	0.451	0.874	4.60	0.033	0.901	8.95

4.1.2 Soil grinding effect on NIR calibrations

Overall, calibration results for SOC and N were not similar with respect to different grinding levels. Measurements in a big cup delivered the best calibrations results for ground samples for SOC (Table 14). The calibration accuracy increased as follows: sieved samples < crushed samples < ground samples. A factorial ANOVA on predicted values of cross validations confirmed a significant difference between ground samples in regard to crushed and sieved samples for SOC. It was not clear which grinding level delivers the best calibration results by scanning samples in a small cup. The RMSECVs and R^2 varied differently for all grinding levels. In general, ground oven-dried samples showed the highest RPIQ. For N crushed samples delivered the best calibration results in 75% of the cases, but not throughout the whole factorial test. The highest RPIQ derived from crushed oven-dried samples. A factorial ANOVA on predicted values out of cross validation showed no significant differences of grinding levels for N.

Table 14: Calibration results for SOC and N of air-dried and oven-dried soil samples with different grinding levels: comparisons between measurements in a small and a big NIR-cup.

	SOC						N					
Air-dried samples												
	RMSECV		R^2		RPIQ		RMSECV		R^2		RPIQ	
	big cup	small cup	big cup	small cup	big cup	small cup	big cup	small cup	big cup	small cup	big cup	small cup
sieved	0.489	0.432	0.736	0.820	2.35	2.81	0.035	0.031	0.894	0.974	4.16	8.21
crushed	0.454	0.486	0.817	0.787	2.79	2.61	0.030	0.030	0.930	0.939	5.05	5.39
ground	0.411	0.449	0.853	0.752	3.13	2.47	0.033	0.035	0.919	0.955	4.74	6.26
Oven-dried samples												
sieved	0.551	0.471	0.573	0.753	1.91	2.42	0.034	0.036	0.936	0.857	5.30	3.63
crushed	0.393	0.397	0.856	0.866	3.15	3.21	0.032	0.032	0.979	0.922	9.32	4.61
ground	0.391	0.452	0.880	0.760	3.45	2.53	0.028	0.034	0.952	0.838	6.09	3.41

But in a comparison between sieved and ground samples, ground samples generally led to smaller error values (RMSECV), except for air-dried samples measured in a small cup: here the lowest error value was shown by crushed samples. Comparing sieved and ground samples in air-dried condition and scanned in a big cup, ground samples delivered the best calibration results (Table 14), whatever the number of repeated measurements (one to four) of sieved samples (Table 13). Ground samples showed the lowest RMSECV, highest R^2 and highest RPIQ values for SOC and N as well.

4.1.3 Effect of cup size on NIR calibrations

No significant differences between cup sizes were recognized for SOC and N; this was confirmed via an ANOVA on predicted values of cross validations. Nevertheless, ground and crushed samples in big cups tended to deliver lower RMESCVs than in small cups (Table 14). Additionally, sieved samples tended to deliver lower RMSECVs in small cups than as measured in a big cup.

4.1.4 Effect of drying on NIR calibrations

If ground and crushed samples were used, calibration accuracy for SOC was generally better for oven-dried samples than for air-dried samples (Table 14). For ground samples, drying led to significantly better results; calculated by an ANOVA on predicted values of cross validations for oven-dried and air-dried ground samples. Figure 11 shows the impact of drying on the shape of soil spectra for two mean spectra (calculated out of 95 soil spectra); one for oven-dried ground and one for air-dried ground samples. Oven-dried samples, scanned in the reflection mode, show a typical decreasing of the trenches of the water absorption between 1400-1900 nm, in comparison to air-dried samples. Moreover, the lowest RMSECV with 0.391 could be reached using oven-dried ground samples. For N it was not clear whether the drying of samples led to better calibration results. R^2 and RMSECVs varied for different grinding levels and cup sizes, independent of drying. However, the lowest RMSECV was delivered by oven-dried ground samples with a value of 0.028.

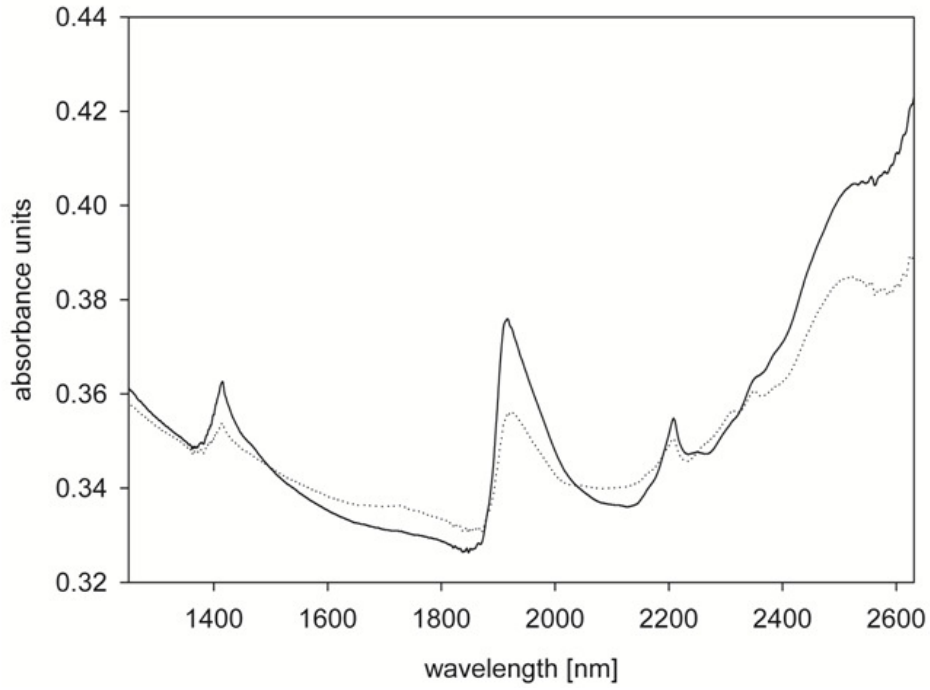

Figure 11: Two mean reflectance spectra, derived from 95 soil samples; air-dried ground samples (solid line), oven-dried ground samples (dotted line).

4.1.5 Factorial effect of grinding, cup size and drying

Ring cup size and drying had little influences on calibration results for SOC and N (Figure 12). An influence of 5% was not reached by one of these sample pretreatments. Nevertheless, the effect of drying was three times higher for SOC than for N. Moreover, the

effect of cup size could be neglected for SOC and the value was close to zero. For N the RMSECV got influenced by the size of the ring cup with close to five percent.

Grinding of samples had the greatest effect on calibration results (Figure 12). The RMSECV was influenced by different grinding levels for SOC with 35% and for N with 28%. Overall, the following order of increasing influence on calibration accuracy could be observed: for SOC cup size < drying < grinding level, and for N drying < cup size < grinding level.

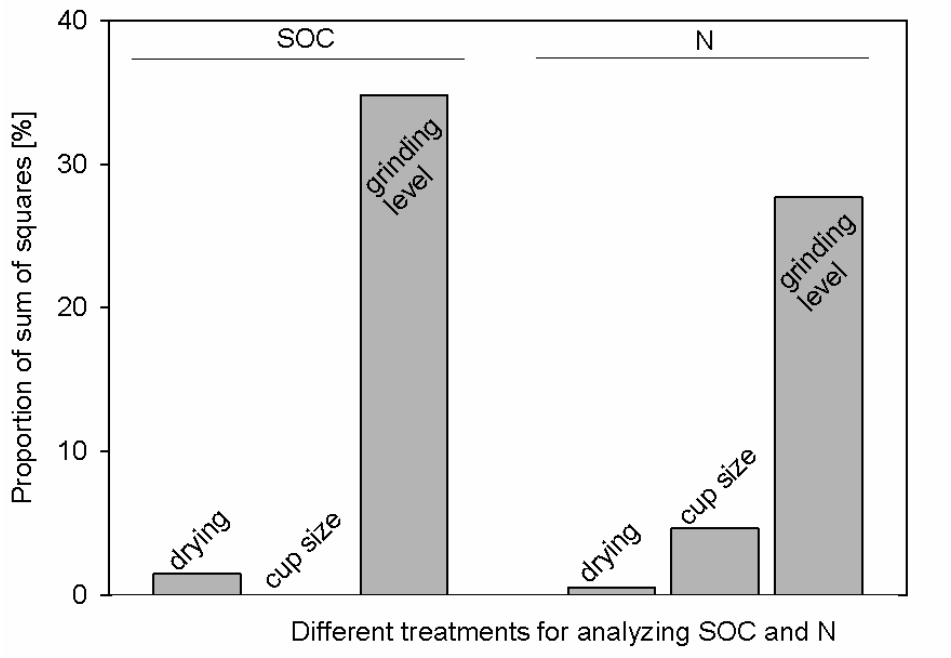

Figure 12: Percentage distribution of sum of squares for different soil treatments, based on a factorial ANOVA on RMSECVs for SOC and N; residuals for SOC hold 63.71% and for N 67.18%.

4.1.6 Soil temperature effect

The NIR spectra were recorded at three different temperatures in order to assess the influence of fluctuating lab temperatures on the spectra and the estimated soil properties. Spectra recorded at 20°C were used for calibration and spectra recorded at 24 and 28°C were used to predict SOC and total N. Prediction accuracy varied considerably for air-dried samples analysed at 24 and 28°C (Table 15). The RMSEP of a 24°C prediction was more than three times higher than the one for a 28°C prediction for SOC. The results for N showed similar results, with a more than two times higher RMSEP for samples recorded at 24°C in comparison to 28°C samples. So, in general, prediction accuracy at 28°C was higher than at 24°C, with lower RMSEP and a higher r^2. Prediction results for oven-dried samples showed a

different pattern. Both temperatures showed a similar accuracy of prediction for both temperatures. Changes of RMSEP were even less than 2% for SOC and N. Factorial ANOVAs underlined no significant differences for oven-dried predictions, but significant differences for air-dried predictions. This temperature experiment showed that oven-dried samples used for calibration and prediction could lead to reproducible RMSEPs, whereas investigations with air-dried samples did not show such a good reproducibility through time.

Table 15: Comparison between prediction results for SOC and N for 24°C and 28°C samples, based on a 20°C calibration.

Calibration set: air-dried samples (20°C)→ RMSECV: SOC=0.393, N=0.029, R^2: SOC=0.917, N=0.926, RPIQ: SOC= 3.97, N=4.89								
Soil temperature	SOC, prediction				N, prediction			
	RMSEP	r^2	RPD	RPIQ	RMSEP	r^2	RPD	RPIQ
24°C	0.962	0.696	0.77	2.06	0.053	0.620	1.16	2.38
28°C	0.290	0.857	2.56	3.28	0.021	0.881	0.94	4.24
Calibration set: Oven-dried samples (20°C), RMSECV: SOC= 0.439, N= 0.033, R^2: SOC= 0.883, N= 0.918, RPIQ: SOC= 2.13, N= 4.52								
24°C	0.636	0.789	1.17	2.06	0.050	0.76	1.2	2.24
28°C	0.639	0.711	1.16	1.77	0.057	0.65	1.90	1.90

4.1.7 Reproducibility of measurements

First, in order to analyse the reproducibility of calibration accuracy, NIR measurements were carried out for three different sample pretreatments. Table 16 shows results of repeated measurements with a time difference of nine months. For N the calibration accuracy varied little for repeated measures within one sample treatment.

This could be seen clearly for the RMSECV. The value varied sparsely for all treatments between 0.027 and 0.035. For SOC, repeated measurements showed similar results, except for ground air-dried samples. Here repeated measures seemed to have a negative effect on calibration results and the RMSECVs differed in a range between 0.461 and 0.478. In this case SOC seemed to be more sensitive towards repeated measures than N. And on average the

RPIQ for nearly all N was also higher than for SOC. However, ground oven-dried samples had the best calibration accuracy with the lowest RMSECV and highest R^2. For all analyses the R^2 varied, since the lowest RMSECV of all calibration possibilities was taken for interpretation and different PLS factors were included.

Table 16: Calibration results of repeated measures with a time distance of nine months: investigated factor was SOC and N for three different sample pretreatments.

Pretreatment	Time of measurement	SOC			N		
		RMSECV	R^2	RPIQ	RMSECV	R^2	RPIQ
ground air-dried	October 2009	0.389	0.879	3.47	0.033	0.906	4.79
	June 2010	0.487	0.831	2.93	0.034	0.860	3.94
sieved air-dried	November 2009	0.478	0.849	2.95	0.035	0.894	4.10
	July 2010	0.461	0.857	3.02	0.033	0.904	4.32
ground oven-dried	November 2009	0.384	0.993	4.68	0.027	0.938	5.89
	July 2010	0.374	0.960	5.95	0.030	0.848	3.79

For a closer look at the reproducibility of SOC and N prediction, which is based on repeated NIR measurements, Figure 13 gives an overview on predicted data for measurements in 2009 and 2010 with a time difference of nine months. Correlation coefficients of predicted values varied between 0.92 and 0.97 for SOC and between 0.96 and 0.97 for N. Air-dried sieved samples delivered the highest correlations of predicted values in comparison to the other pretreatments for both properties. For ground samples, oven-dried samples showed a higher correlation of SOC (R^2=0.94) during time than air-dried samples (R^2=0.92) (Figure 13).

Second, when air-dried samples were used for calibration and prediction sample moisture is poorly controlled. Throughout the literature no standard protocols for sample drying exist, questioning the comparability of different soil sample NIR scans (Russell, 2003; He et al., 2005; Brunet et al., 2007). In order to analyse the reproducibility of NIR measurements of not standardised dried samples, we recorded spectra of air-dried samples for a later calibration as well as spectra of samples that were recorded in different time intervals after drying for prediction. Table 17 shows a wide variation of RMSEPs and r^2 for SOC as well as for N, independent of time.

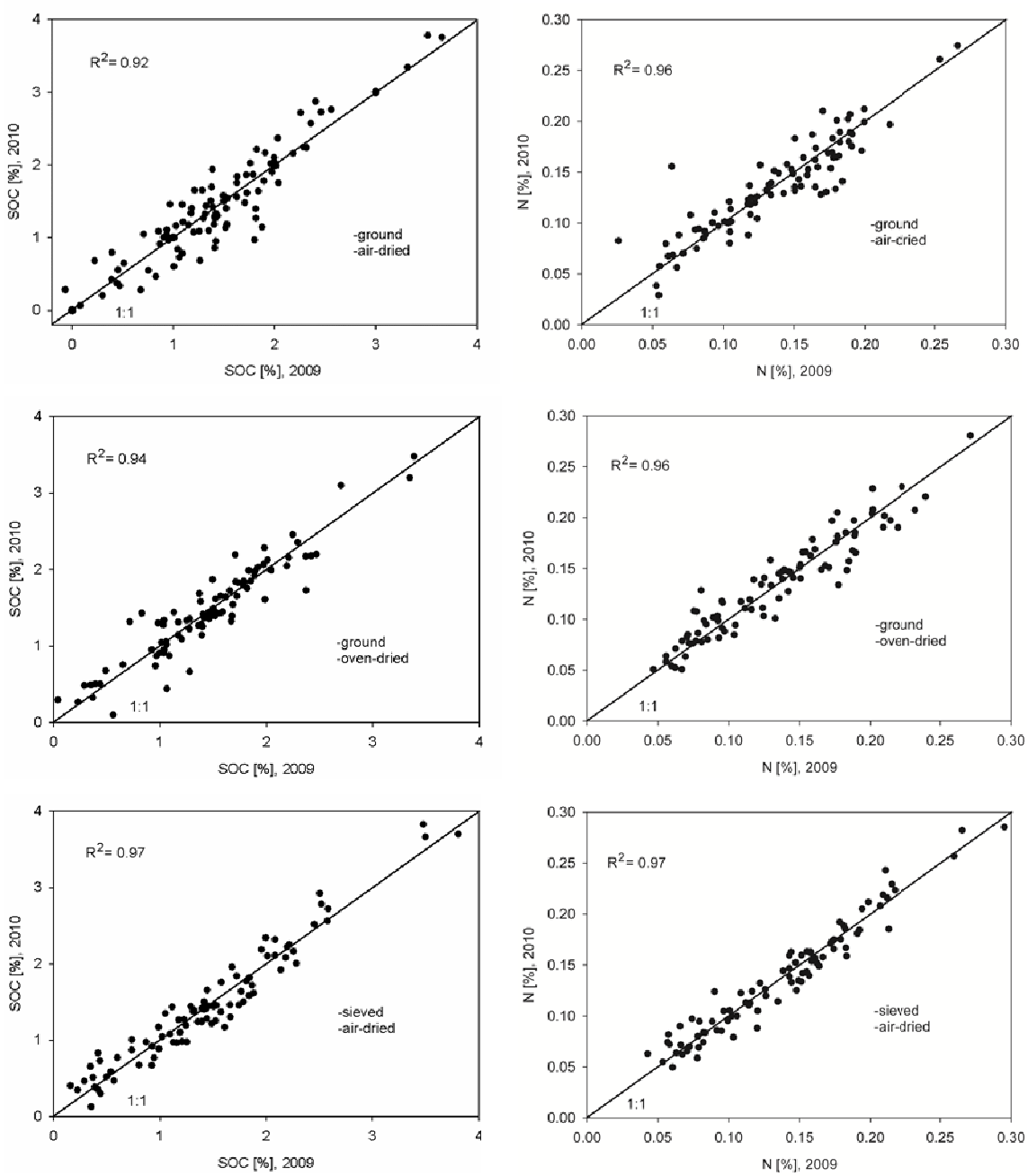

Figure 13: Correlations of predicted values out of best calibration models for SOC and N; measurements were repeated with a time distance of nine months between 2009 and 2010.

Table 17: Prediction results of dried and not standardised dried samples for SOC and N, based on a calibration of air-dried samples, with '*' = standardised drying and '°' = not standardised drying.

Calibration set: air-dried samples, R^2: C= 0.949, N= 0.928, RMSECV: C= 0.392, N= 0.032, RPIQ: C= 5.15 , N= 4.97								
Time of measurement after drying	C, prediction				N, prediction			
	RMSEP	r^2	RPD	RPIQ	RMSEP	r^2	RPD	RPIQ
directly after drying *	0.981	0.857	0.74	2.29	0.033	0.838	1.81	3.06
1 week after drying °	0.297	0.842	2.44	3.15	0.022	0.868	2.70	4.09
2 weeks after drying °	0.304	0.842	2.38	3.18	0.022	0.876	2.73	4.18
5 weeks after drying °	0.568	0.788	1.27	2.79	0.025	0.841	2.43	3.67
9 months after drying °	0.377	0.742	1.88	2.51	0.031	0.686	1.76	2.87
9 months later directly after drying *	0.756	0.546	0.94	1.30	0.038	0.699	1.45	2.36

A closer look on predicted values confirmed the variation of predicted SOC and N for not standardised dried samples (Figure 14). The median of all air-dried samples ranged between 1.37 and 1.83 % SOC and between 1.13 and 1.15% N. Samples that were recorded directly after drying should serve as a general comparison to air-dried samples. Values of oven-dried samples seemed to be over predicted and the median for SOC was around 2.14 and for N around 0.16. In addition, repetition of measurements in oven-dried condition seemed to lead to reproducible results, whereas predictions for air-dried samples varied with time.

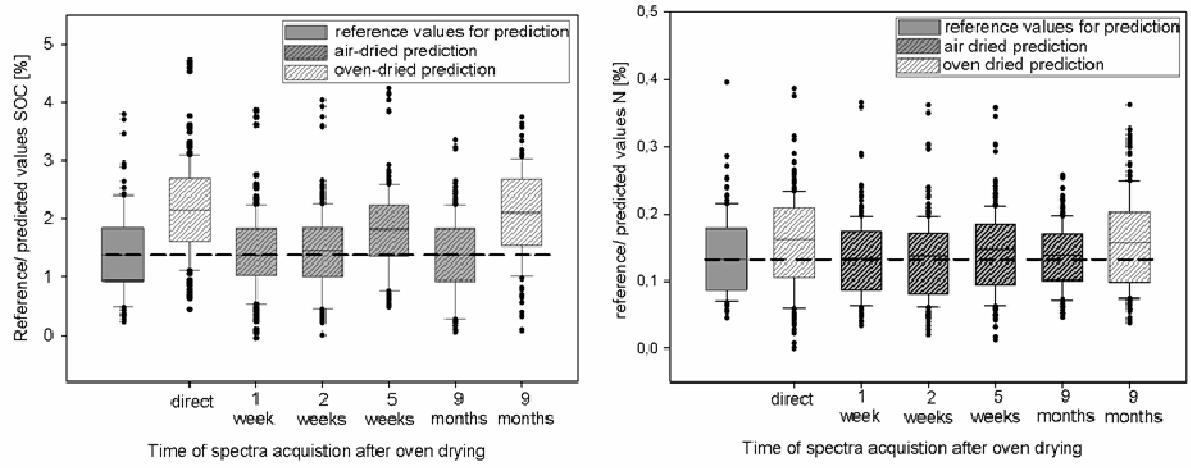

Figure 14: Variation of predicted SOC and N for oven-dried and air-dried samples, based on an air-dried calibration.

4.2 Field-based NIR measurements

Absorption spectra scanned via horizontal or vertical NIR measurements were obviously different for the three fields (Figure 15). Between 1200 and 2000 nm, the distinct water peaks appeared at 1400 and 1900 nm (Russell, 2003; Miltz and Don, 2012). However, the peak area differed according to field and measurement method. Baseline offsets (position of spectral baseline) differed from the lowest to highest absorption: FALNO < HdBA < Espenberg (for shank and probe examinations). The absorption at 450 nm showed the presence of iron oxide minerals (i.e. goethite and haematite), which could be seen in all soil spectra (Clark, 1999) recorded via probe analysis.

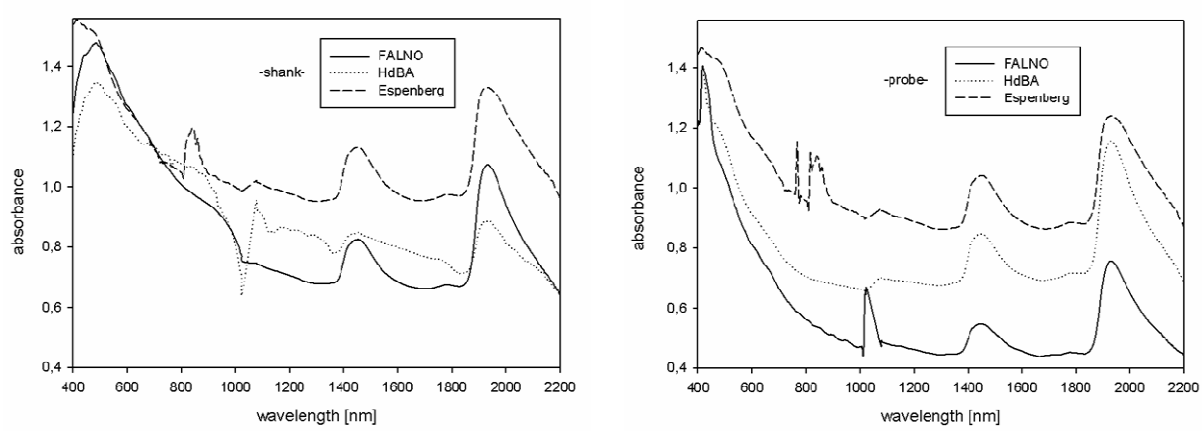

Figure 15: Average NIR spectra of the FALNO, HdBA and Espenberg fields scanned via on-line shank (left) and probe (right) measurements. Spectra from reference soil samples, representing the 0–10 cm depth, served for averaging (shank: FALNO = 36 reference spectra, HdBA = 18 reference spectra and Espenberg = 12 reference spectra; probe: 12 reference spectra per field, spectra were averaged spectra representing the 0–10 cm depth).

Quartz, the dominant mineral in sandy soils, was not found in any of the topsoil spectra, with a typical peak at around 600 nm (Russell, 2003). Carbonate minerals with a peak at around 1930 nm (Craig et al., 2009) could not explicitly be found since the wide water peak at around 2000 nm overlapped smaller peaks in this region. The transition zone between 1000–1100 nm should not be interpreted since here, the visible region ends at around 1040 nm, and the NIR region started at around 1080 nm, depending on the spectrometer calibration. The small gap of about 40 nm was not closed by mathematical calculations from the VERIS software.

4.2.1 Horizontal field mapping: shank

Overall, there were 545 spectra ha^{-1} recorded on FALNO, 775 spectra ha^{-1} measured on HdBA and on Espenberg, 199 spectra were collected per ha while driving 11 km h^{-1} (HdBA), 12 km h^{-1} (FALNO) and 20 km h^{-1} (Espenberg) on average. The driven speed had an influence on the number of collected soil spectra: The lower the tractor speed of tractor, the more spectra could be gained per area. The high speed driven on Espenberg was the outcome of the field's altitude, with a hilltop and two slopes of different levels: A short distance uphill could only be managed very slowly (because of the limit of the tractor's traction force), and a longer distance downhill could be managed quickly. However, a total of 20 metres elevation difference had to be overcome on that field.

Table 18: Results of shank and probe measurements per field investigation on FALNO, HdBA and Espenberg: recorded field spectra, cleaned field spectra for calibration and the number of removed outlier spectra due to a default Mahalanobis distance.

Fields	Shank				Probe			
	Field spectra	Cleaned field spectra	Number of removed outlier spectra	Removed spectra [%]	Field spectra	Cleaned field spectra	Number of removed outlier spectra	Removed spectra [%]
FALNO	5015	4807	208	4.15	3379	2994	385	11.39
HdBA	6509	6288	221	3.39	2038	1433	605	29.69
Espenberg	2305	2305	0	0	1669	763	906	54.28

On FALNO and HdBA, the shank was pulled in a sinuous line through the field, without lifting the plough out of the soil at the field boundary, to ensure a continuous measurement. On Espenberg, the tractor also drove in parallel lines over the field – always from north to south – with several measurement interruptions at the field's boundaries due to its altitude difference. The measurement results that could be used for further processing were the so-called cleaned field spectra (Table 18) that simply represented all scanned field spectra without outliers – these were sorted out according to their Mahalanobis distance.

Table 19: Calibration results of SOC and total N for the fields FALNO, HdBA and Espenberg, investigated via shank on-line NIRS and VIS-NIRS.

Fields	SOC			N		
	NIR					
	R^2	RMSECV	Pretreatment	R^2	RMSECV	Pretreatment
FALNO	0.96	0.172	SNV	0.99	0.016	SNV
HdBA	0.91	0.061	SNV	0.98	0.011	SNV
Espenberg	0.90	0.149	SNV	0.94	0.014	SNV
	VIS-NIR					
FALNO	0.73	0.162	SNV	0.97	0.016	refl
HdBA	0.82	0.060	1^{st} der	0.80	0.006	SNV
Espenberg	0.64	0.218	SNV	0.64	0.021	SNV

The number of collected reference soil samples differed per field (Table 4). However, the number of reference samples was in the range with those of Christy (2008) and Knadel et al. (2011), who successfully calibrated the C content out of 15 soil samples. Different soil sampling methods had varying expenditures of time: The random one could be carried out at the fastest rate (it took five minutes to reach a field position and sample), followed by the cluster technique. This last method implied a calculating process using the VERIS software which was followed by a soil sampling at calculated field positions (the calculation time was dependent on the number of field spectra for processing: 2–10 minutes was normal).

The calibration results were divided into two parts: results calculated i) out of NIR spectra (1100–2000 nm) and ii) out of VIS-NIR spectra (400–2200 nm) (Table 19). The NIR calibrations of all three fields were successful, with a $R^2 \geq 0.9$ for SOC and N when spectra were pretreated with SNV.

On the subject of SOC and N, R^2 increased in the following field order: Espenberg < HdBA < FALNO. The RMSECV was generally not higher than 0.172 for SOC as well as 0.016 for N. The RMSECV decreased in the following field order: FALNO > Espenberg > HdBA.

A different pattern could be seen for VIS-NIR calibrations. All R^2 values here (for SOC and N and for all fields) were lower than R^2 for NIR calibrations. The R^2 went down to 0.64 (lowest value) for SOC and N, and it increased in the following field order: Espenberg < FALNO < HdBA for SOC and Espenberg < HdBA < FALNO for N. It was eye-catching that Espenberg showed the lowest R^2 for all measurements for SOC and N. The RMSECV was not larger than 0.218 for SOC and 0.021 for N. The RMSECV decreased for SOC and N in the following order: HdBA <FALNO <Espenberg. Here, SNV was the most

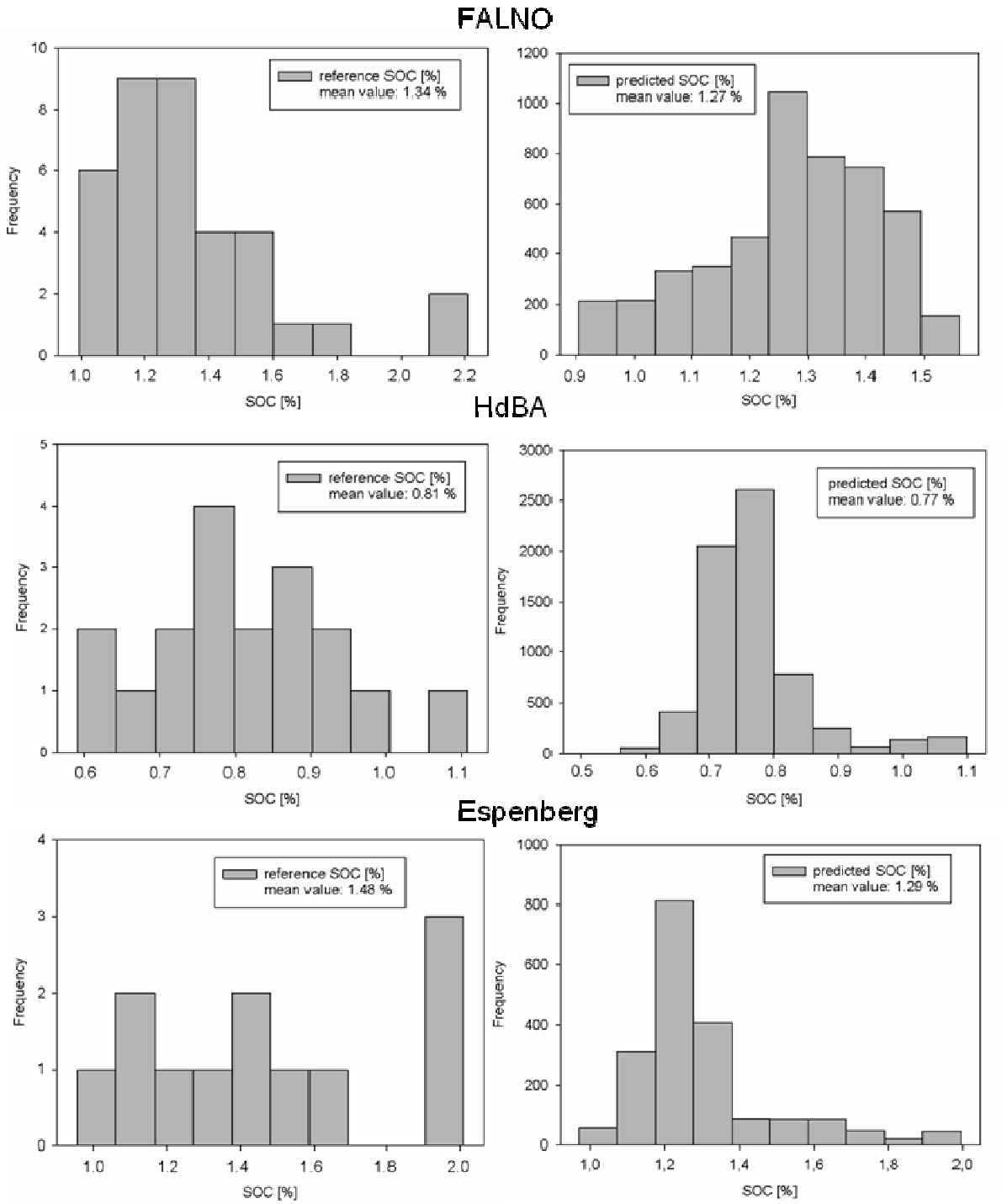

Figure 16: Frequency distribution of SOC concentrations of soil samples measured via direct chemical analysis for reference purposes or predicted (using the best calibration model) on the basis of on-line scanned NIR spectra. Field measurements were carried out with NIR shank on the fields FALNO, HdBA and Espenberg.

applied spectra manipulation (4 of 6 cases), but reflectance spectra as well as the first derivative also led to the best calibration results. For all measurement series, six spectra manipulations were performed (see section 3.2.4). The use of different manipulations, such as the ones listed below, always led to R^2 lower than 0.5. Overall, NIR calibrations led to higher

R^2 than did VIS-NIR calibrations. On the basis of the best calibration models (all from NIR investigations with satisfying calibration results), SOC concentrations were predicted, needed for the later SOC stock calculation. Both, the reference and predicted SOC concentrations for the fields were displayed in Figure 16. The range of predicted SOC values was mostly smaller than the range of reference SOC data and the mean values of predicted SOC concentrations decreased in all cases, compared to the reference mean values.

4.2.2 Vertical field mapping: probe

Logging soil real-time data via the probe gave a vertical picture of soil characteristics. The maximum soil depth was 70 cm for all three fields, depending on the increasing penetration resistance with depth that was encountered and the cumulative friction on the probe rod (see section 3.2.2.2). Cleaned field spectra (for all depths) – which were of value for further calculations – were 763 for Espenberg, 1433 for HdBA and 2994 for FALNO (Table 18). Regarding SOC, calibration results yielded in $R^2 \geq 0.90$ for FALNO and HdBA (Table 20). The RMSECV varied between 0.025 and 0.154 for FALNO and between 0.021 and 0.086 for HdBA. Universal calibrations (including all depth sections) revealed $R^2 \geq 0.91$ for both fields. Here, the RMSECVs were larger than for single-depth calibrations. In general, Espenberg showed worse calibration results, with lower R^2 and higher RMSECV for single-depth and universal calibration results in comparison to the other fields. Here, the R^2 varied in a wide range between 0.42 and 0.92 for single-depth calibrations, and the RMSECV ranged between 0.077 and 0.532. On the whole, spectral manipulations varied per field and per depth. Most pretreatments were non-modified original reflectance spectra (seven times) and those of first derivative (six times). However, there is no recognisable trend for spectra manipulations for certain single-depth sections, whereas universal calibrations always revealed the best calibration results for SNV-treated spectra.

Similar results were obtained for N-calibrations: FALNO and HdBA showed higher R^2 (single-depth and universal calibration: $R^2 \geq 0.91$) and lower calibration errors (single-depth calibration: RMSECV ≤ 0.011, universal calibration: RMSECV ≤ 0.028) than did the results for Espenberg. R^2 for Espenberg ranged between 0.49 and 0.97 for single-depth calibrations, and the RMSECV was up to 0.336 for N. Here, the universal calibration resulted in $R^2 = 0.72$ and RMSECV = 0.025. Spectral manipulations for N-calibrations were similar to those of SOC-calibrations: first derivative (six times) and reflectance spectra (six times) were the most occurring spectra manipulations. For universal calibration, the SNV spectra led to the best calibration results for N.

Table 20: Calibration results for SOC and total N for the fields FALNO, HdBA and Espenberg, investigated via probe VIS-NIR spectroscopy.

Depth [cm]	SOC								
	FALNO			HdBA			Espenberg		
	R^2	RMSECV	Pretreatment	R^2	RMSECV	Pretreatment	R^2	RMSECV	Pretreatment
0-10	0.90	0.129	SNV	0.99	0.065	1st der	0.66	0.248	refl
10-20	0.92	0.147	SNV	0.95	0.072	1st der	0.66	0.158	refl
20-30	0.96	0.154	normal	0.98	0.086	refl	0.69	0.424	2nd der
30-40	0.98	0.085	1st der	0.99	0.021	refl	0.92	0.365	refl
40-50	0.94	0.044	detrend	0.92	0.027	refl	0.70	0.077	1st der
50-60	0.97	0.025	detrend	0.99	0.034	1st der	0.61	0.129	refl
60-70	0.90	0.041	detrend	0.99	0.030	1st der	0.42	0.532	normal
0-70	0.92	0.232	SNV	0.91	0.104	SNV	0.48	0.438	SNV
N									
0-10	0.96	0.011	1st der	0.99	0.008	1st der	0.61	0.028	refl
10-20	0.94	0.006	1st der	0.99	0.003	refl	0.60	0.016	refl
20-30	0.98	0.005	SNV	0.98	0.009	refl	0.49	0.336	2nd der
30-40	0.93	0.010	detrend	0.97	0.002	refl	0.97	0.021	1st der
40-50	0.92	0.004	detrend	0.99	0.005	refl	0.95	0.014	2nd der
50-60	0.98	0.002	detrend	0.99	0.004	1st der	0.54	0.024	1st der
60-70	0.91	0.003	detrend	0.99	0.001	2nd der	0.91	0.009	SNV
0-70	0.94	0.028	SNV	0.91	0.009	SNV	0.72	0.025	SNV

4.2.3 Cross-checking of shank and probe

4.2.3.1 Data interpolation

The application of Kriging was successful, generating SOC stock semivariograms out of SOC stock data predicted from spectra that were obtained via on-line shank NIR measurements. Out of eight possible data interpolation models, two models fitted best: Exponential without nugget was the best for FALNO and Espenberg, and the M. Stein parameterisation with nugget fitted best for HdBA. The best model fits were similar for both SOC stock types, for SOC stock CALL and SOC stock FIELDSPEC (see Table 21).

Table 21: Kriging results for SOC stocks FIELDSPEC and CALL based on predictions from spectra scanned via on-line shank NIR measurements: The best regression models were selected via AIC criterion, leading to sill, range, nugget values and nugget to sill ratios for the three fields studied.

Fields	Best regression model	SOC stock FIELDSPEC (SOC stock model A)			
		Sill [(Mg ha^{-1} 30 cm^{-1})^2]	Range [m]	Nugget [(Mg ha^{-1} 30 cm^{-1})^2]	Nugget to sill ratio [%]
FALNO	Exponential without nugget	51	55.1	-	-
HdBA	M. Stein with nugget	52.8	51	6.3	10
Espenberg	Exponential without nugget	20.1	25.3	-	-
		SOC stock CALL (SOC stock model B)			
FALNO	Exponential without nugget	56.6	60.5	-	-
HdBA	M. Stein with nugget	84.8	81	12.5	20
Espenberg	Exponential without nugget	28.3	32.5	-	-

For example, Figure 17 illustrates a semivariogram of SOC stock data for HdBA (SOC stock model A) with M. Stein parameterisation. Spatial dependencies among neighbouring calculated SOC stocks showed a correlation of up to a 51 m distance. For FALNO, field points correlated even up to a 55.1 m distance, and for Espenberg, the spatial correlation of SOC stocks was lower, with a range of 25.3 m, both values referring to SOC stock model A (see Table 6). The nugget-to-sill ratio for HdBA was 10%. Such a low nugget-to-sill ratio indicated a small degree of small-scale variation. The semivariograms for FALNO and Espenberg did not have a nugget effect. Small errors in the measurement value as well as enough data can lead to a low or no nugget effect, respectively. Sill and range values for SOC stock CALL data (SOC stock model B) were consistently larger than for SOC stock FIELDSPEC data with regard to all fields. For probe examinations, Kriging could not be adequately applied, and semivariograms showed pure nugget.

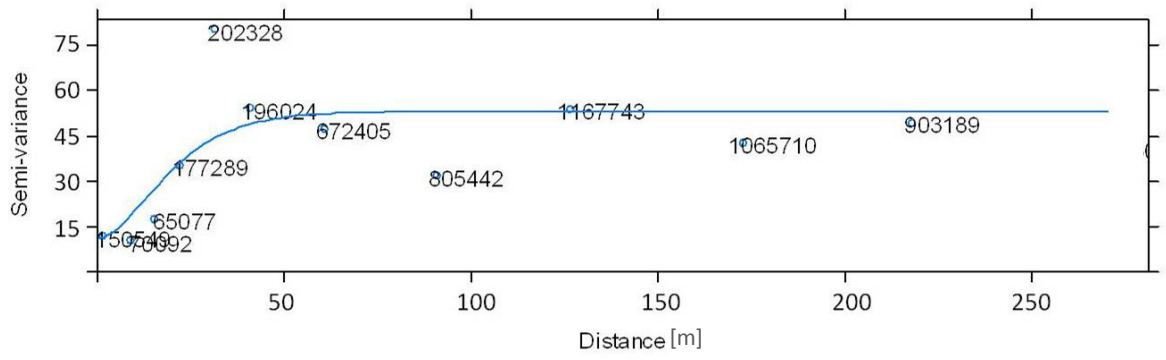

Figure 17: Semivariogram of SOC stock FIELDSPEC data (SOC stock model A) of the HdBA field.

4.2.3.2 Calculation of SOC stocks (0–30 cm depth) and SOC stock error

Eight different model options for SOC stock assessment were computed out of SOC data generated by direct chemical analysis in the laboratory and out of BD data estimated by different PTFs (Table 22). SOC stocks varied between 58.1 and 63.4 Mg ha^{-1} 30 cm^{-1} for FALNO, between 38.0 and 42.3 Mg ha^{-1} 30 cm^{-1} for HdBA and between 61.7 and 74.1 Mg ha^{-1} 30 cm^{-1} for Espenberg. In seven of eight model options, SOC stocks showed the following increasing order: HdBA < FALNO < Espenberg. Additionally, SOC stocks CALL were larger than SOC stocks FIELDSPEC for the most part. Here, SOC stock models with a similar calculation base were compared with each other (model H with model G; model J with model I; model M with model L).

Table 22: SOC stocks for FALNO, HdBA and Espenberg computed out of eight different calculation models (G–N), with standard error (SE_{REF}) and measurement error (SE_{MEAS}) added. Original SOC data were from laboratory analysis (direct chemical analysis).

Field	SOC stock model	Type of soil samples	SOC stock [Mg ha^{-1} 30 cm^{-1}]	SE_{REF} [Mg ha^{-1} 30 cm^{-1}]	SE_{MEAS} [Mg ha^{-1} 30 cm^{-1}]
FALNO	G	composite	58.9	1.85	1.76
	H		59.1	7.42	7.24
	I	soil core	59.6	1.24	1.22
	J		58.4	10.91	6.68
	K		58.1	4.01	1.44
	L		63.4	2.35	2.08
	M		61.0	7.49	7.43
	N		60.9	2.36	2.07
HdBA	G	composite	41.9	1.37	1.08
	H		42.3	3.28	3.23
	I	soil core	39.8	1.42	1.12
	J		39.9	4.88	3.29
	K		40.3	1.42	1.11
	L		39.1	1.25	1.10
	M		38.0	3.91	3.25
	N		39.1	1.25	1.08
Espenberg	G	composite	69.8	4.02	2.20
	H		74.1	4.96	2.58
	I	soil core	61.7	6.99	1.88
	J		62.5	6.61	5.29
	K		61.8	7.11	1.76
	L		73.5	4.59	2.21
	M		74.1	6.65	6.06
	N		73.9	4.51	2.10

SOC stock models I, J and K (data available for 0–30 cm) calculated for FALNO and Espenberg led to smaller SOC stocks than did comparable stock models L, M and N (data available for 0–10 cm), which is probably due to large SOC concentrations in the upper 0–10 cm depth section extended to a 0–30 cm depth with similar SOC data (models L, M and N). In contrast, SOC stocks models L, M and N calculated for HdBA were smaller than for SOC stocks I, J and K. This is most likely associated with the large SOC concentrations in the 10–20 cm depth section considered solely in SOC stock models I, J and K. To distinguish between and to evaluate different SOC stock models, the standard error (SE_{REF}) and measurement error (SE_{MEAS}) were calculated. The SE_{MEAS} is directly incorporated with the SE_{REF} so that the SE_{MEAS} is always smaller than the SE_{REF}, as shown in Table 22. The SE_{MEAS} was 2.45 to 99.19% of the SE_{REF} for FALNO, 0.11 to 98.37% for HdBA and 1.57 to 91.23% for Espenberg. SOC stocks calculated out of BD_{CALL} regressions led to the highest SE_{MEAS} in all cases (models H, J and M). Here, in five of six cases, the measurement error was lower for shank than for probe investigations.

Table 23: SOC stocks for FALNO, HdBA and Espenberg computed out of six different calculation models (A–F), with prediction errors SE_{PRED} and SE_{RMSEP}. SOC $stock_{MC}$ was added with its standard error (SE_{MC}). Predicted SOC data for SOC stock assessment based on NIR field measurements.

Field	SOC stock model	Type of soil samples	SOC stock [Mg ha^{-1} 30 cm^{-1}]	SE_{PRED}	SE_{RMSEP}	SOC $stock_{MC}$ [Mg ha^{-1} 30 cm^{-1}]	SE_{MC}
FALNO	A	composite	57.3	0.38	0.08	58.5	3.77
	B		54.9	0.40	0.07	58.2	2.66
	C	soil core	62.4	5.50	0.34	58.6	8.84
	D		61.1	5.61	0.64	59.5	9.49
	E		65.7	3.09	0.26	71.3	8.95
	F		63.8	3.01	0.35	70.2	9.58
HdBA	A	composite	42.3	0.20	0.03	41.5	2.7
	B		49.2	0.53	0.04	47.0	7.15
	C	soil core	41.7	1.11	0.12	41.8	2.15
	D		42.7	1.27	0.25	42.8	2.46
	E		44.5	0.87	0.16	44.7	1.89
	F		45.9	1.01	0.30	46.1	2.20
Espenberg	A	composite	58.4	0.08	0.02	57.3	1.46
	B		58.2	0.08	0.09	52.9	1.97
	C	soil core	58.6	0.86	0.11	61.9	4.79
	D		59.5	1.30	0.28	60.6	7.24
	E		71.4	0.80	0.08	65.7	6.00
	F		70.1	1.01	0.29	63.7	7.39

Regarding SOC stocks computed out of BD_{FS} regressions, the measurement error was lower for shank measurements in 50% of the cases. With regard to SOC stocks computed out of on-line NIR measurements, there is no recognisable trend for SOC stock values across all models and fields (Table 23). With respect to error values and as mentioned above, SE_{PRED} is the equivalent standard error to the SE_{REF}, but it included additional measurements with NIRS. Shank measurements (SOC stock models A and B) led to low SE_{PRED}, in contrast to consistently larger SE_{PRED} when probe measurements were included (SOC stock modes C–F). Most SE_{PRED} calculated with BD_{CALL} (SOC stock models B, D, F) held larger errors than calculated without it (SOC stock models A, C, E) in eight of nine cases. In a comparison between shank and probe measurements and SOC data utilisation from the 0–10 cm depth, the SE_{PRED} of SOC stocks generated out of probe-investigations was larger (model E > model A; model F > model B) than that computed out of shank measurements. With regard to the error of SOC stocks based on on-line NIR measurements (SOC stock models A–F), the SE_{RMSEP} was smaller than the SE_{PRED} in 17 of 18 cases: The SE_{RMSEP} values were up to one-tenth of the SE_{PRED} values. A Monte Carlo simulation error SE_{MC} was generally larger than SE_{RMSEP} or SE_{PRED}. It was around 10 times larger than the SE_{PRED} and around 50–150 times larger than the SE_{RMSEP}. In addition, when BD_{CALL} regressions were used for SOC stock estimation (SOC stock models B, D, F), the SE_{MC} was larger in eight of nine cases in comparison to results including BD_{FS} regressions (SOC stock models A, C, E). In a comparison between the use of shank or probe, the SE_{MC} was larger for SOC stock estimation via probe than via shank in eight of 12 cases. Only SOC stocks from HdBA held consistently larger SE_{MC} when on-line shank NIR measurements were performed. In a comparison, errors from SOC stock assessments with on-line NIR application (SE_{MC}) were smaller than those from direct chemical analysis (SE_{REF}) in around 60% of the cases (Figure 18).

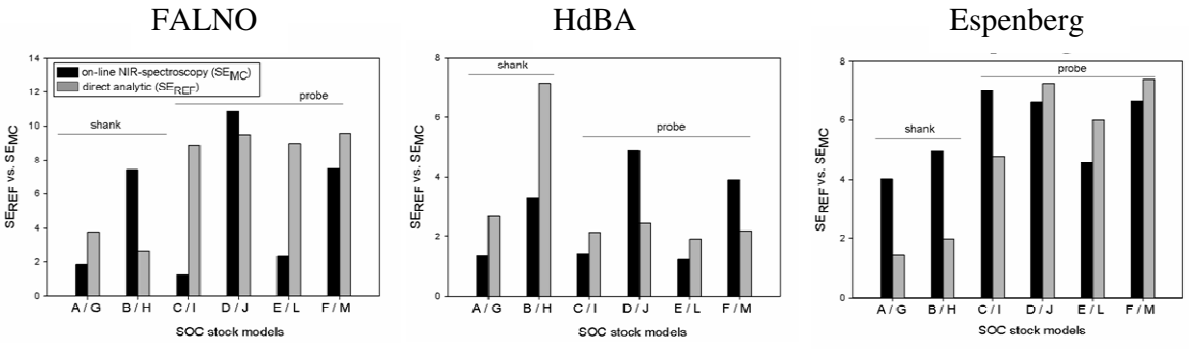

Figure 18: Comparison between SOC stock errors of different SOC stock models A–M that arose from SOC stock estimation via direct chemical analysis of SOC concentration (SE_{REF}) or via an additional use of on-line NIRS with shank and probe application (SE_{MC}).

4.2.3.3 Time aspect of SOC stock estimations

In general, SOC stock estimation via on-line NIR measurements took more time than that via the sole use of direct chemical analysis. This can be seen in Table 8, where field and laboratory work are listed according to all individual steps and their amount of time. Laboratory work comprised 6.9 hours for SOC stock estimation (direct chemical analysis) and the field work covered 11.75 hours. Since field and laboratory work were combined for SOC stock assessment including on-line NIR measurements, the overall working time was even longer in this case.

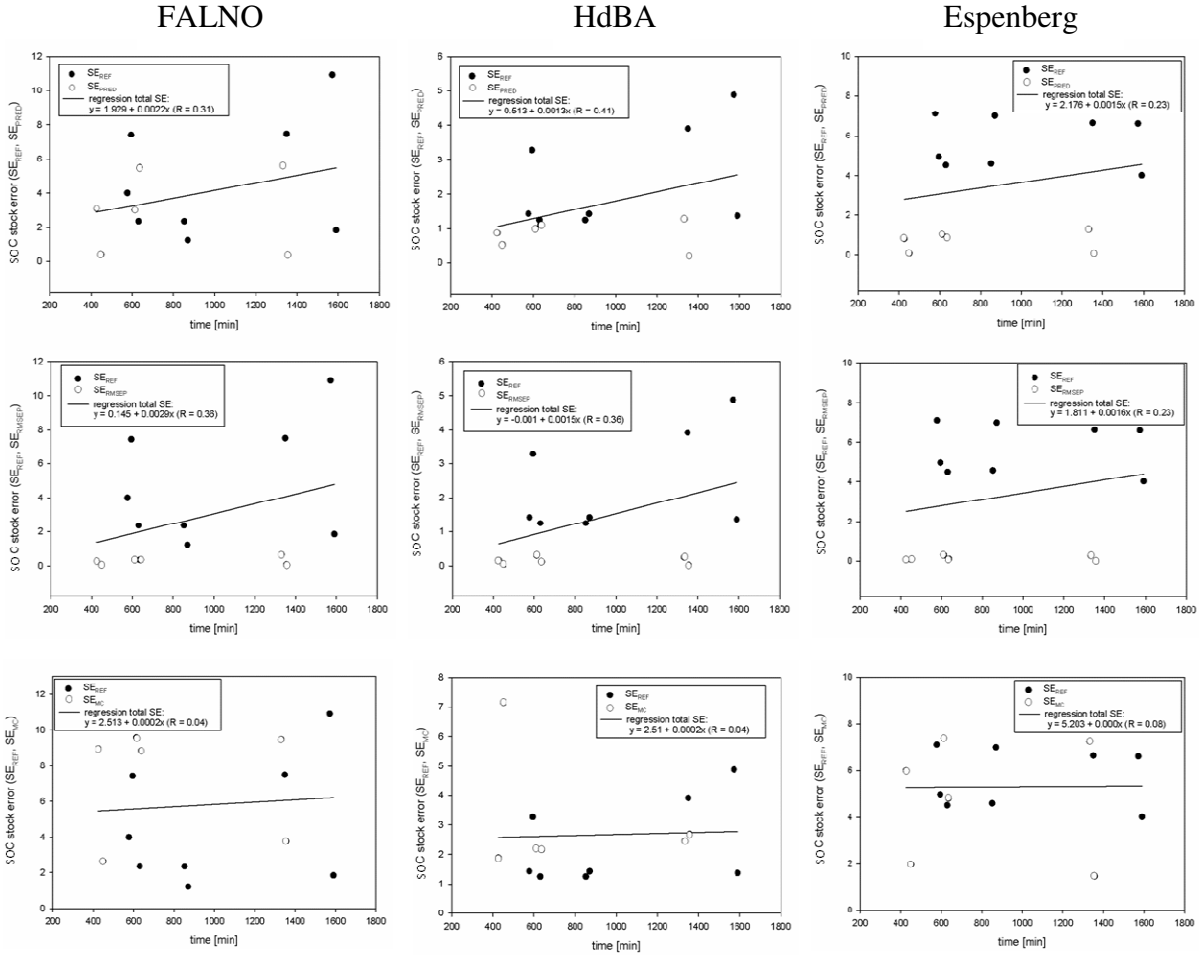

Figure 19: Overview of time versus SOC stock error for FALNO, HdBA and Espenberg. SOC stock errors are SE_{REF} (direct chemical analysis, light data points) and SE_{PRED}, SE_{RMSEP} and SE_{MC} (on-line NIR measurements, dark data points).

Figure 19 presents an overview of SOC stock error versus time for SOC stock estimation. Here, the SOC stock error SE_{REF} (direct chemical analysis) was faced with three SOC stock errors that included on-line NIR measurements: SE_{PRED}, SE_{RMSEP} and SE_{MC}. Overall, correlations between time and different SOC stock errors were relatively low, with $R \leq 0.41$.

Comparatively, the strongest correlations emerged when SE_{REF} was plotted with SE_{PRED} or SE_{RMSEP} ($0.23 \leq R \leq 0.41$). The smallest correlations were identified when plotting SE_{REF} with SE_{MC} against time. However, a slight rise of SOC stock error with time could generally be recognised. Furthermore, the working time and costs that were needed for SOC stock estimation were calculated. In general, field and laboratory work can be managed by well-trained student- and technical assistants. They handle over more than 75% of the total amount of work (Tables 8 and 24).

The rest is accomplished by the research assistant. The total costs for SOC stock assessment based on shank with additional probe measurements amounted to 1424.65 €, including 30.65 hours working time. The costs and working hours were allocated in the following order: student assistant < research assistant < technical assistant. Regarding the use of shank and probe separately, 12.95 hours and costs of 636.15 € are needed for SOC stock estimation via shank measurements, and 17.7 hours and 788.50 € are needed when probe measurements are included. The difference between a shank and probe based SOC stock assessment was not large. However, the use of shank measurements would be generally quicker (working time is reduced by ~27%) and cheaper (costs are reduced by ~20%) for SOC stock determination.

Table 24: Outline of the students' (STA), technical assistants' (TA) and research assistants' (RA) gross wages and working time for the three required workspaces in order to assess SOC stocks on the basis of shank and probe NIR measurements. Data in Table 8 serve as a basis for calculation.

Investigation via	Shank			Probe		
Workspace	Working time [h]	Operator	Gross wage	Working time [h]	Operator	Gross wage
Field	3	TA	135.30 €	4	TA	180.40 €
	0.75	STA	21.45 €	4	STA	114.40 €
Laboratory	6	TA	270.60 €	6	TA	270.60 €
	0.2	STA	5.72 €	0.7	STA	20.02 €
Office	3	RA	203.08 €	3	RA	203.08 €

4.3 Field reproducibility study

4.3.1 Soil NIR spectra as well as calibration results for SOC

The soil spectra of FAL250 and HdBA can be seen in Figure 20. Here, field NIR spectra recorded at reference samplings positions were averaged into one field spectrum to get a short overview on the fields' spectral characteristics. All spectra had two absorption peaks at 1400–1600 cm^{-1} and at 1850–2150 cm^{-1} in common. However, the shape of spectra from repeated measurements differed little in the maximum of absorption peaks as well as in the offset of the baselines.

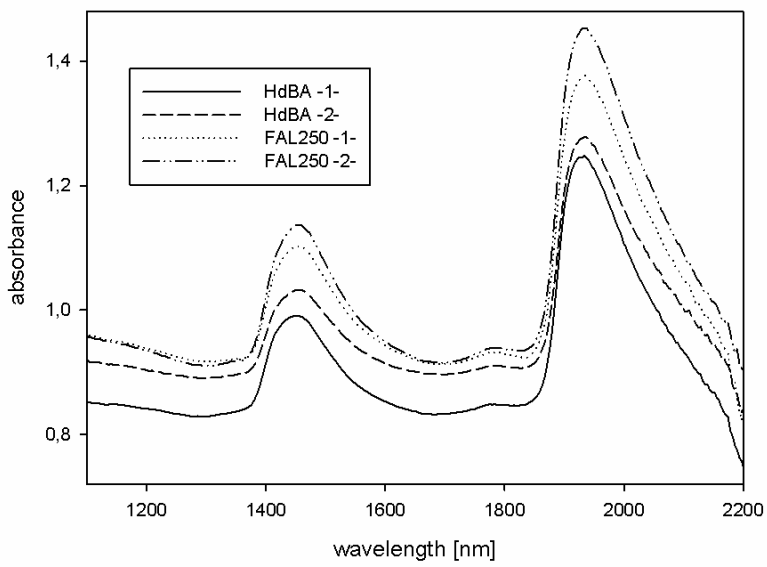

Figure 20: Average NIR spectra of the fields HdBA and FAL250 from repeated measurement campaigns. Spectra from reference sampling positions served for averaging (HdBA = 18 sampling positions and spectra, FAL250 = 19 sampling positions and spectra.

Table 25: SOC calibration results for the fields FAL250 and HdBA from repeated measurement campaigns. The number of PLSR components, removed outliers and type of spectra pretreatment for best calibration accuracy were added. RMSECV values are displayed in g kg^{-1} SOC; this SOC unit was used throughout the "field reproducibility study".

Field	Measure-ment	Number of ref. spectra and SOC conc. for calibration	SOC							
			R^2	RMSECV	SEC	RPD	RER	PLSR compo-nents	Spectra pretreat-ment	Out-lier
FAL250	-1-	19	0.93	1.03	6.5	1.74	7.8	8	refl	0
	-2- ^{rep}	19	0.90	1.31	8.3	1.37	6.1	9	refl	0
	-2-	19	0.93	1.28	7.9	1.41	5.3	10	refl	0
HdBA	-1-	17	0.99	0.93	10.3	1.40	5.6	8	1^{st} der	1
	-2-	17	0.99	0.39	3.1	1.46	23.4	7	refl	1

The overall calibration results for SOC yielded in $R^2 \geq 0.90$, RMSECV ≤ 1.31, SEC ≤ 10.3, RPD ≥ 1.37 and RER ≥ 5.3 for repeated measurements on FAL250 and HdBA (Table 24). Overall, the studies on HdBA led to higher R^2 and lower RMSECV than the studies on FAL250 (Table 25). With respect to spectra manipulation, the reflectance spectra led to the best calibration results in four of five cases, and once, the first derivative achieved best calibration accuracy. The number of PLSR components ranged from seven to 10, generally. Furthermore, for HdBA, one spectral outlier was determined throughout both calibrations of repeated measurements, having a strong leverage effect on the calibration (measurement one: H-outlier, measurement two: D-outlier, both outliers refer to different soil samples and spectra). Therefore, the outlier was removed before the final calibration model was fitted. This approach allowed for the reduction of the influence of potential outlier data points that could be caused by errors during SOC analysis and NIR measurements (Sorensen and Dalsgaard, 2005). The worst calibration results were obtained from measurement -2-^{rep} of FAL250, with the lowest R^2 and RPD-value as well as the highest RMSECV-value. The highest standard error of calibration of all FAL250 measurements was attained by measurement -2-^{rep}, also.

4.3.2 Results of SOC concentrations and SOC stocks: direct chemical analysis versus on-line NIR measurements

Overall, the average SOC concentrations out of the reference analysis varied between 14.7 and 15 g kg^{-1} for FAL250 and between 8.1 and 9.3 g kg^{-1} for HdBA as results derived from repeated measurement campaigns (Table 26). The coefficient of variation (CV) of SOC was greater for the first measurement, with 16%, and lower for the second measurement campaign, with 6% for HdBA. The CV values for SOC of FAL250 were 12% in all measurement cases. The predicted average SOC concentrations ranged between 15.8 and 16.1 g kg^{-1} within repeated measurement campaigns on FAL250 and between 9.2 and 9.5 g kg^{-1} on HdBA.

All average values were calculated out of SOC concentrations from off-line predictions. Here, the number of predicted SOC concentrations was equal to the amount of field spectra scanned during each measurement campaign, with the exception of measurement one on HdBA. The average value of 9.3 g kg^{-1} was based on 1830 predicted SOC concentrations (instead of 1870 predicted SOC concentrations from 1870 cleaned field spectra), since 40 negative SOC concentrations were received through the off-line prediction being useless for further processing. Moreover, the largest CV values for the predicted SOC content, with 16% and

17%, arose from the measurements on FAL250. For HdBA, the CV values of the predicted SOC content were very small, with 1% and 2%. The CV values of SOC stocks calculated out of the reference analysis were between 8% and 10% for FAL250 and between 6% and 15% for HdBA. In general, the CV values of SOC stock FIELDSPEC were equal or slightly smaller than those of SOC stock CALL. The difference of CV values was 2% for FAL250 and 9% for HdBA in respect to repeated measurement campaigns, independently of SOC stock type. The difference of CV values of predicted SOC stocks out of repeated measurement campaigns was smaller, with 5% for HdBA and no change in SOC variation could even be detected for FAL250. For HdBA, all values for SOC and SOC stocks were larger when they resulted from the second rather than from the first measurement campaign, consistently. This was in line with the reference data of FAL250 (-1- and -2-). With regard to predicted values, the data from SOC concentrations and SOC stocks FIELDSPEC were larger from the first and lower from the second measurement campaign (-2-^{rep} and -2-). Moreover, repeated measurements on FAL250 (low time interval of two weeks) yielded lower SEs for SOC concentrations and SOC stocks, when NIR spectra and SOC data were both scanned and analysed a second time (-2-).

Table 26: Averaged SOC concentrations and SOC stocks measured and calculated out of results from direct chemical analysis (REF) or out of on-line NIR measurements (TEST) for the fields FAL250 and HdBA from repeated measurement campaigns. SOC stock data derived from a field-specific regression (FIELDSPEC) and from a Callesen regression for sandy soils (CALL). The second column shows the standard error (SE) and the third column the coefficient of variation (CV) of SOC concentrations or SOC stocks.

Field	Measure-ment	REF (basis: direct chemical analysis)								
		SOC [g kg^{-1}]			SOC stock FIELDSPEC [Mg ha^{-1} 30 cm^{-1}]			SOC stock CALL [Mg ha^{-1} 30 cm^{-1}]		
		mean	SE	CV	mean	SE	CV	mean	SE	CV
FAL250	-1-	14.7	0.5	12	66.58	1.85	10	64.41	1.93	11
	-2- ^{rep}	14.7	0.5	12	66.58	1.85	10	64.41	1.93	11
	-2-	15	0.4	12	70.13	1.39	8	68.05	1.44	9
HdBA	-1-	8.1	0.3	16	38.40	1.37	15	36.45	1.32	15
	-2-	9.3	0.2	6	43.77	0.73	6	41.55	0.71	6
TEST (basis: on-line NIR measurements)										
FAL250	-1-	16.1	0.05	17	71.99	0.22	15	70.24	0.23	15
	-2- ^{rep}	15.9	0.05	16	71.66	0.23	15	69.65	0.23	15
	-2-	15.8	0.01	16	71.21	0.20	15	69.14	0.21	15
HdBA	-1-	9.2	<0.01	1	41.03	0.09	14	38.77	0.07	14
	-2-	9.5	<0.01	2	44.11	0.09	9	43.05	0.07	9

The difference in the SEs of SOC and SOC stocks from the second measurement campaign (-2-^{rep} and -2-) were smaller for the predicted values and larger for the reference values.
In addition, averaged SOC concentrations and SOC stocks were generally larger when they derived from on-line NIR measurements and lower when measured and calculated solely in the laboratory via direct chemical analysis. In addition, the reference and predicted SOC stocks were consistently larger when a field-specific regression was used, and lower values were achieved for calculations with a Callesen regression.

4.3.3 Semivariogram results of SOC concentrations and SOC stocks

The statistical information of a field can be presented by the mean value of the measured parameter. Additional information on the spatial distribution and dependencies between measured values can be obtained by the semivariance. In order to study the variation of SOC concentrations and SOC stocks quantitatively, semivariograms were generated, giving three spatial dependency parameters: sill, range and nugget. The function that fitted best to all semivariograms was an exponential fit, with and without nugget. Tested spherical, Gaussian and M. Stein's models always led to high AIC values with a high variance of the residues. The semivariograms of the reference SOC data were of special interest since these values served as a basis for the calibrations and calculations of SOC stocks. Nevertheless, these semivariograms showed pure nugget or an unreliable combination of an exponential function in a low lag distance, followed by a horizontal line at a higher lag distance as model fit. These fits were therefore not used for interpretation. However, the semivariograms of predicted SOC data showed acceptable model fits for FAL250 and HdBA from the first and second measurement campaigns (Figure 21).
The statistical data of semivariograms can be seen in Table 27. Here, the differences of range and sill for repeated measurements for FAL250 were low. The sill of the SOC concentrations ranged from 0.63 to 0.79, and the range varied between 36 m and 45 m. For HdBA, the sill of SOC concentration amounted to 0.153 $(g\ kg^{-1})^2$ (first measurement campaign) or to 0.0014 $(g\ kg^{-1})^2$ (second measurement campaign) and the range to 64 m or 74 m from the first and second measurement campaigns, respectively. For FAL250, the sill-values were larger and the range-values smaller than corresponding data from HdBA in all measurement cases. Repeated measurement campaigns for estimating the SOC stock FIELDSPEC of FAL250 showed similar values for range, sill and nugget. In contrast, HdBA had different values for sill, range and nugget. On the subject of SOC stock CALL, values for sill and range were generally

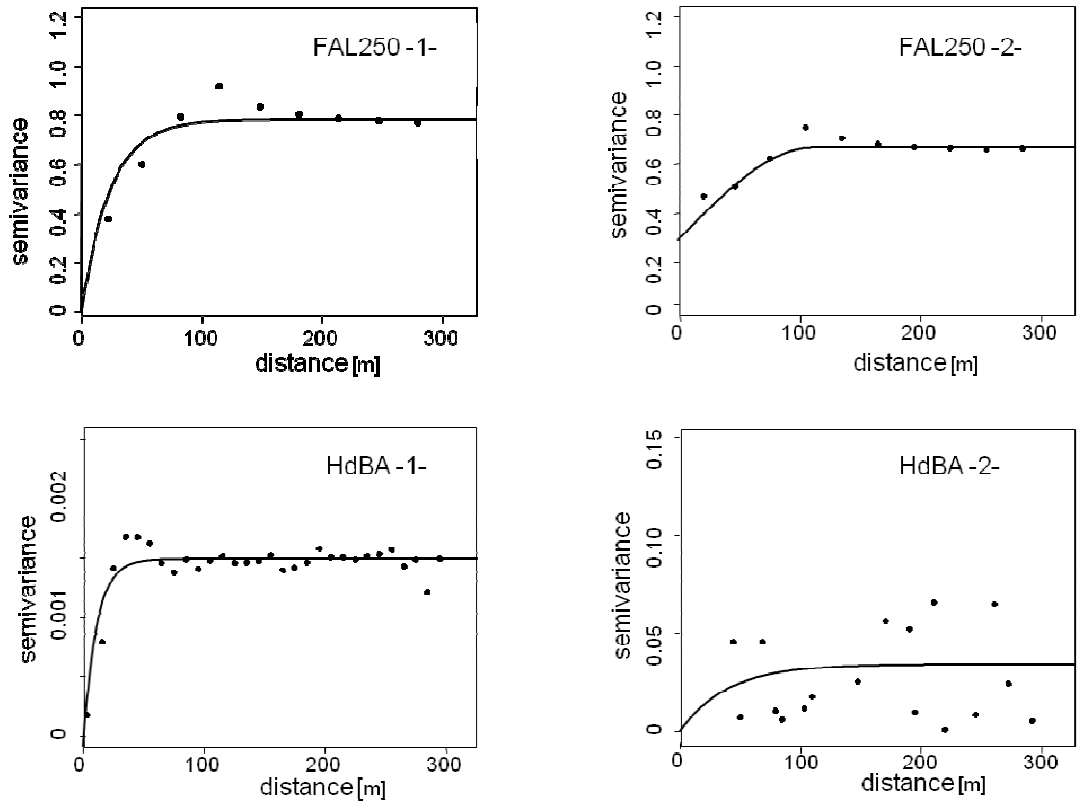

Figure 21: Semivariograms of predicted SOC concentrations for repeated measurement campaigns on FAL250 and HdBA. The basis was a calibration with reference SOC data and a following prediction of SOC data from field spectra collected via on-line NIRS. Best model fits were exponential fits solely.

larger than for SOC stocks FIELDSPEC and for SOC concentrations for both investigated fields. For FAL250, the following range order could be recognised: SOC concentrations < SOC stock FIELDSPEC < SOC stock CALL. The sequence changed in regard to both measurements on HdBA: SOC stock FIELDSPEC < SOC concentrations < SOC stock CALL. Measurement -2- ^{rep} from FAL250 showed sill, range and nugget values that were in between the value-range from measurements one and two (-1- and -2-), except the semivariograms of SOC concentration and SOC stock CALL, both having no nugget effect (Table 27). Overall and for repeated measurement campaigns, the maximum difference of estimated SOC concentration ranges was 9 m for FAL250 and 10 m for HdBA.

Considering SOC stocks, the maximum range difference on FAL250 was 2 m for SOC stock FIELDSPEC and 6 m for SOC stock CALL. HdBA had a range difference of 13 m for SOC stock FIELDSPEC and 10 m for SOC stock CALL.

Table 27: Parameter of semivariograms for predicted SOC concentrations [g kg^{-1}] and SOC stocks [Mg ha^{-1} 30 cm^{-1}], estimated with exponential model fits, with and without nugget, for the fields FAL250 and HdBA and repeated measurement campaigns. The SOC concentrations and SOC stocks used for semivariogram estimation derived from calibration of reference data with a following prediction of NIR field spectra. Sill and nugget values for SOC concentration are displayed in g kg^{-1}; sill and nugget values for SOC stocks are shown in (Mg ha^{-1} $30cm^{-1}$)^2.

Field	Measure-ment	Semivariogram parameter	SOC concentration	SOC stock FIELDSPEC	SOC stock CALL
FAL250	-1-	Sill	0.79	36.59	70.26
		Range [m]	36	46	49
		Nugget	-	34.6	-
		Nugget/sill [%]	-	95	-
	-2- rep	Sill	0.68	36.31	75.21
		Range [m]	40	46	52
		Nugget	-	32.12	-
		Nugget/sill [%]	-	89	-
	-2-	Sill	0.63	37.34	81.71
		Range [m]	45	48	55
		Nugget	0.29	31.53	-
		Nugget/sill [%]	46	84	-
HdBA	-1-	Sill	0.0014	52.8	84.77
		Range [m]	64	51	81
		Nugget	0.001	6.3	12.5
		Nugget/sill [%]	71	10	20
	-2-	Sill	0.153	64.05	65.01
		Range [m]	74	38	91
		Nugget	-	-	-
		Nugget/sill [%]	-	-	-

4.3.4 Kriging prediction

Kriging is a data interpolation method that integrates the spatial autocorrelation of the regression residuals of a relevant soil component into the Kriging system. The ordinary Kriging maps of SOC concentrations from the on-line NIR measurements of FAL250 and HdBA can be seen in Figures 22 and 23. In general, darker colours represented higher soil SOC concentrations, and lighter colours denoted lower SOC concentrations. Difference maps were a result of subtracting the SOC data of the first from the SOC data of the second measurement campaign.

Both measurement campaigns of HdBA delivered Kriging maps with different SOC distributions in the same SOC concentration range between 8 (light grey) and 10 g kg^{-1} (dark grey) on average (Figure 22). However, what both Kriging maps had in common were zones with an average SOC concentration range between 9.8 and 10 g kg^{-1} in the same area: two dark grey zones on the eastern part of the field for the second measurement and four dark grey zones for the first investigation at the same place. Moreover, there were two small regions in the eastern part of both measurements that had the lowest average SOC concentration of the field with 8 g kg^{-1} (light grey). The difference between both Kriging maps was as follows: There were few areas with different SOC concentrations on map two (~ low variability) and many small areas with different SOC concentrations for map one (~ high variability). The difference in SOC concentrations from both measurement campaigns varied between 0.2 and 0.8 g kg^{-1} on average (Figure 22). The Kriging-derived maps of SOC concentrations from the first and second measurement campaigns for FAL250 can be seen in Figure 23. The SOC pattern in the field was similar for repeated measurement campaigns. Regions in the north and

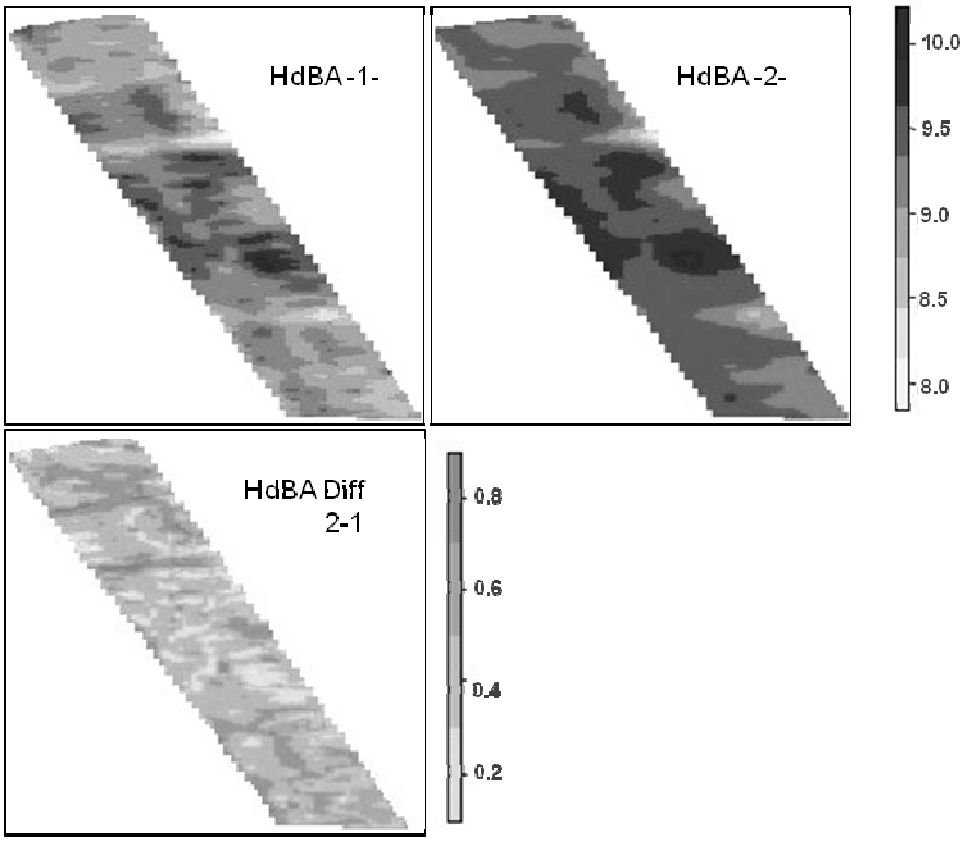

Figure 22: Field HdBA – Ordinary Kriging of SOC concentration (g kg^{-1}) with 6288 data points from the first and 1582 data points from the second on-line NIR measurement campaign. The third Kriging picture displays the calculated difference of SOC concentrations from the second and first measurement campaign.

middle of the field had a SOC concentration of 16 and 18 g kg^{-1} on average (dark grey). The SOC concentration of the soil became lower in the southern direction and went down to a concentration of 9.5 g kg^{-1} on average. The difference of SOC concentrations from both measurement campaigns varied between 0 to 2 g kg^{-1} on average (Figure 23). The zone with no SOC difference occupied ~80% of the whole field area.

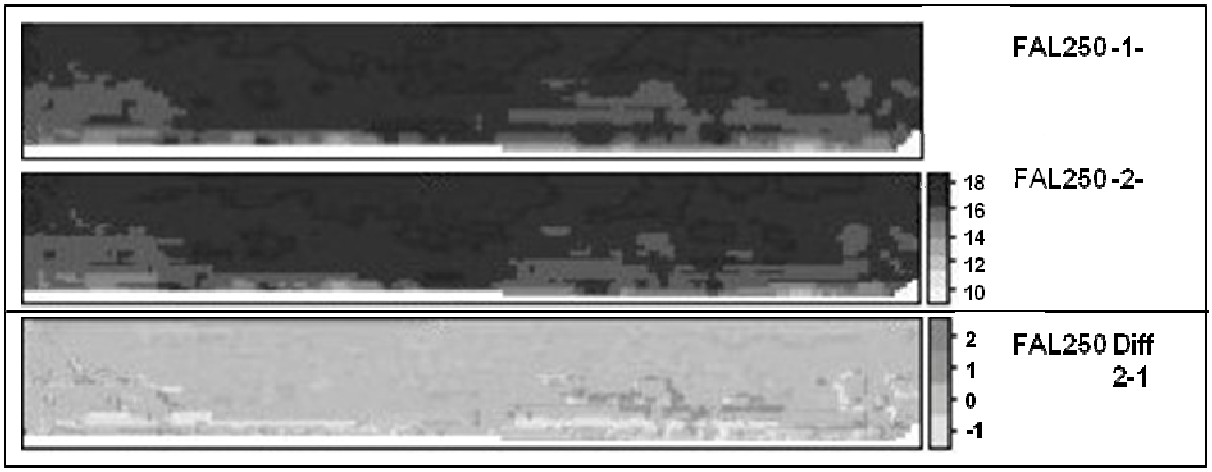

Figure 23: Field FAL250 – Ordinary Kriging of SOC concentration (g kg^{-1}) with 1227 data points from the first and 2136 data points from the second on-line NIR measurement campaign. The third Kriging picture displays the calculated difference of SOC concentrations from the second and first measurement campaign.

4.3.5 MDD of SOC concentrations and SOC stocks

The MDD was calculated to check the effectiveness of direct chemical anaylsis versus on-line NIRS. Within this inquiry, the MDD was estimated twice for repeated measurement campaigns in order to test its reproducibility. For FAL250, the MDD was estimated a third time, combining SOC concentrations from the first measurement campaign with the NIR spectra from the second measurement campaign for calibration. Overall, changes in SOC concentration between 0.2 and 2.1 g kg^{-1} as well as SOC stock changes between 1.3 and 8.1 Mg ha^{-1} 30cm^{-1} were found to be theoretically detectable (Table 28). Moreover, the MDD was lower for predicted (TEST) than for reference SOC concentrations and SOC stock types (REF), independent of field and investigation time. The MDD values for measurement -2-^{rep} of FAL250 were the lowest in all measurement cases from FAL250, regarding the SOC concentrations and both SOC stock types.

Overall, the difference in the MDDs of SOC concentrations and SOC stocks from repeated measurement campaigns was lower for predicted values and larger for all reference

computations. The repeated measurement campaigns of HdBA revealed a detectable SOC concentration difference of 7% for reference and of even 0% for predicted SOC data (Table 28). Here, the difference in MDDs from reference SOC stocks was larger, with 8% for both SOC stock types, and lower for predicted SOC stocks, with 1% for SOC stock FIELDSPEC and even no change for SOC stock CALL.

Table 28: Minimum detectable difference (MDD) of SOC concentration [g kg^{-1}] and SOC stocks [M ha^{-1} 30 cm^{-1}] based on reference (REF: direct chemical analysis) or predicted (TEST: on-line NIR measurements) data for FAL250 and HdBA, repeated measurement campaigns and two types of SOC stock calculation. The basis of SOC stock calculation was a Callesen regression (CALL) and a field-specific one (FIELDSPEC). Relative MDD was the absolute MDD related to the mean of SOC concentrations or SOC stocks.

Field	Measure-ment	MDD					
		REF			TEST		
		SOC [g kg^{-1}]	SOC stock FIELDSPEC [Mg ha^{-1} 30cm^{-1}]	SOC stock CALL [Mg ha^{-} 30cm^{-1}]	SOC [g kg^{-1}]	SOC stock FIELDSPEC [Mg ha^{-1} 30cm^{-1}]	SOC stock CALL [Mg ha^{-1} 30cm^{-1}]
FAL250	-1-	2.1	7.7	8.1	0.7	3.4	3.7
	-2-^{rep}	1.5	5.5	5.7	0.5	2.5	2.7
	-2-	1.7	5.8	6.0	0.6	3.1	3.6
HdBA	-1-	1.3	5.7	5.5	0.2	1.5	1.4
	-2-	0.8	3.1	3.0	0.2	1.3	1.5
Rel. MDD							
Field	Measure-ment	REF			TEST		
		SOC [%]	SOC stock FIELDSPEC [%]	SOC stock CALL [%]	SOC [%]	SOC stock FIELDSPEC [%]	SOC stock CALL [%]
FAL250	-1-	14	12	13	4	5	5
	-2-^{rep}	10	8	9	3	4	4
	-2-	11	8	9	4	4	5
HdBA	-1-	16	15	15	2	4	4
	-2-	9	7	7	2	3	4

With respect to FAL250, the difference in estimated MDDs from repeated measurement campaigns (-1- and -2-) was 3% for SOC concentration and 4% for both SOC stock types, regarding the reference calculations. The corresponding predicted values showed a lower difference of MDDs for SOC stock FIELDSPEC of 1% and of even 0% for SOC concentration and SOC stock CALL. MDD differences between measurement type two (-2-^{rep}) and those from the first measurement campaign were consistently larger when compared with the MDD differences between the first (-1-) and second (-2-) measurement

campaign. Considering relative MDDs, the differences between measurement type two and the first measurement campaign were larger in 50% of the cases than the difference between the first and second measurement campaigns.

For calculated MDDs of SOC stock FIELDSPEC, it was not relevant whether measurement one was being compared with measurement campaign two or measurement type two (-2- ^{rep}); the MDD differences were equal for both reference and predicted calculations. For SOC stock CALL, the difference in MDDs from measurement two and -2-^{rep} from repeated measurement campaigns was 1% and 0% for reference and predicted computations, respectively.

Figure 24 shows the relationship between the MDD of SOC concentration with soil sample number and soil spectra. Here, the original SOC concentrations were taken for MDD estimation, since they delivered the only data estimated via direct chemical analysis without any further functions for calculation. In contrast, BD values for SOC stock assessment could only be estimated via PTFs, which do not cover the real field variability of soil BD. Curve diagrams for SOC concentration (see Figure 24) and SOC stock were therefore similar, with different values for the maximum MDD. The MDD declined with a rising number of soil

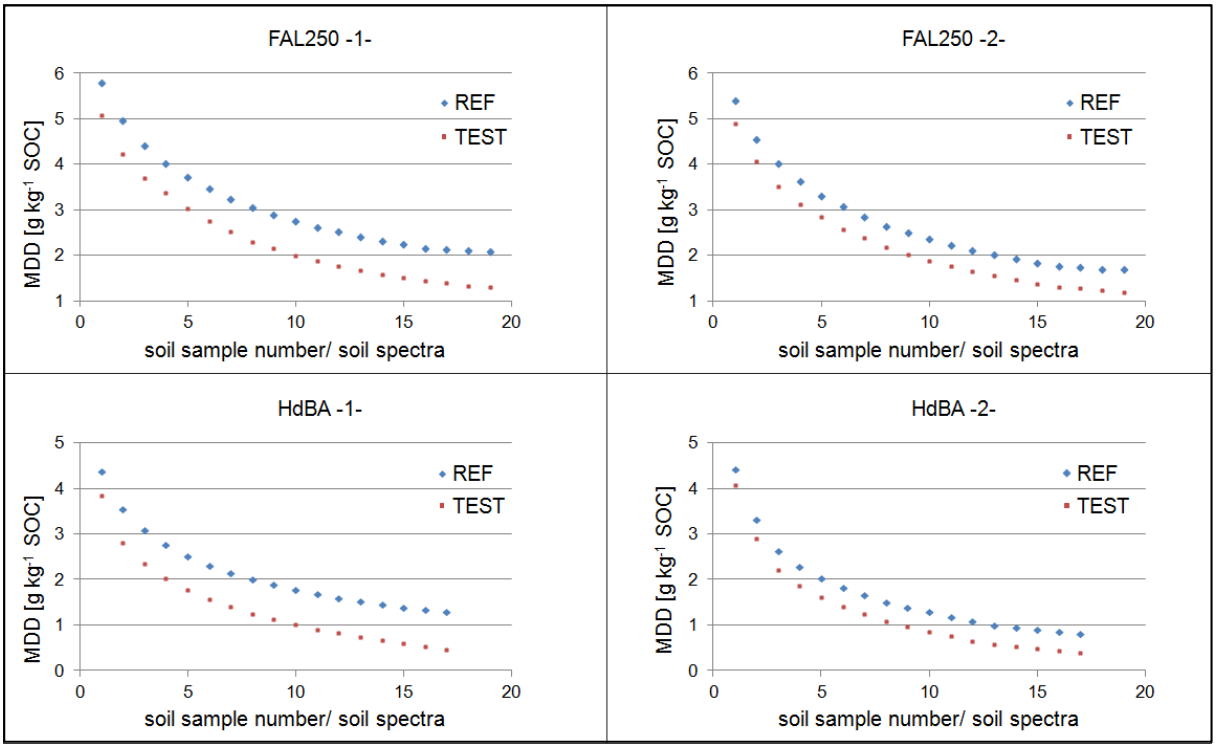

Figure 24: Minimum detectable difference (MDD) of SOC concentration for a 0–30 cm depth and repeated measurement campaigns on HdBA and FAL250 in relation to the number of soil samples and soil spectra taken for SOC calibration; MDDs were compared for reference (REF) and predicted (TEST) SOC concentrations; MDD values were calculated via a Monte Carlo simulation that varied the number of soil samples/ soil spectra per field and measurement campaign.

samples (Figure 24). When the number of samples was increased from two to 10, the MDD dropped significantly, by ≥50% in all cases. When more samples were taken, the MDDs continued to get smaller, but the difference was smaller the more samples that were implicated. For inquiries on FAL250 and HdBA, the MDD of reference SOC concentrations was generally larger than the MDD for predicted concentrations per soil sample number. The soil sample number for reference analysis was restricted by the number of reference soil data taken for SOC calibration, with 17 for HdBA and 19 for FAL250. The number of soil spectra for TEST analysis was even higher. Here, the more samples that were taken for MDD calculation, the MDD continued to lower, until it reached the average MDD value per field and measurement campaign (Table 28).

Having calculated the MDD of SOC stocks and keeping in mind the European annual mean SOC stock loss of 0.17 Mg ha^{-1} yr^{-1}, as modelled by Ciais et al. (2010), the number of years until a SOC stock change can be detected was calculated (Table 29).

When field measurements were repeated, a minimum of eight years was theoretically needed until a change in SOC stocks would be detectable (Table 29), according to the presented investigation methods. Regarding the second measurement campaign type, called -2-^{rep}, for FAL250, it was no longer being considered, since for SOC stock change detection and a minimum time difference between repeated measurements, two separate soil sampling campaigns are necessary. For the sake of completeness, this measurement type is listed in Table 29. In general, 18 to 48 years should pass before a second measurement campaign on FAL250 is undertaken for SOC stock change detection. For HdBA, 8 to 34 years are needed until a second measurement campaign is conducted on HdBA.

However, the time difference of repeated measurement campaigns for SOC stock change detection i) was consistently shorter when on-line NIR measurements (TEST) were conducted instead of solely conventional soil sampling (REF) and ii) was shorter when a field-specific regression for BD-determination was used in comparison to the one from Callesen – in five of eight cases. Furthermore, the reduction of years as the time difference for repeated measurements was consistently more than 80% via the use of on-line NIRS in comparison to direct chemical analysis.

Table 29: Theoretical number of years until a change of SOC stock could be detected. The calculation based on SOC stocks generated out of reference SOC data from lab (REF) or predicted SOC data out of on-line NIR measurements (TEST) from fields FAL250 and HdBA. MDD and annual decrease of SOC stocks of 0.17 Mg ha^{-1} yr^{-1} (Ciais et al., 2010) were used to estimate the number of years. This was done for repeated measurement campaigns. SOC stock calculations were computed via regressions from Callesen et al. (2003) (CALL) and field-specific regressions (FIELDSPEC). The number of SOC data (n) for SOC stock calculation was added.

Field	Measure-ment	Years until SOC stock change detection			
		REF		TEST	
		SOC stock FIELDSPEC	SOC stock CALL	SOC stock FIELDSPEC	SOC stock CALL
FAL250	-1-	45 (n=19)	48 (n=19)	20 (n=80)	22 (n=75)
	-2-^{rep}	32 (n=19)	34 (n=19)	15 (n=140)	16 (n=123)
	-2-	34 (n=19)	35 (n=19)	18 (n=134)	21 (n=117)
HdBA	-1-	34 (n=17)	33 (n=17)	9 (n=370)	8 (n=233)
	-2-	18 (n=17)	18 (n=17)	8 (n=125)	9 (n=52)

5 Discussion

5.1 Laboratory-based NIR measurements

5.1.1 Effect of repeated measures on NIR calibrations

The instantaneous reproducibility of NIR scans from soils samples was high. Thus, the effect of repeated NIR measurements on the calibration error was negligible. Double scans are very common in soil NIR analysis (Fystro, 2002; Russell, 2003; Rinnan and Rinnan, 2007; Terhoeven-Urselmans et al., 2008; Zornoza et al., 2008) and some studies used even more than two scans (Chang et al., 2001; Brunet et al., 2007). Our results showed that recording one instead of two scans per sample had no effect on the calibration quality and could be an effective measure to optimize soil NIR analytics. This is in line with Sorensen and Dalsgaard (2005) who noticed that the accuracy of NIR predictions could not be improved significantly by performing double scans instead of single scans. Also Barthes et al. (2006) found no significant effect of up to six repeated scans on the calibration error for SOC and N of sieved soil samples.

5.1.2 Soil grinding effect on NIR calibrations

Mostly, NIRS measurements for SOC and N are carried out on sieved soil samples (Fystro, 2002; Cozzolino and Moron, 2003; Russell, 2003; He et al., 2005) and less on crushed and ground samples (Chang and Laird, 2002; Brunet et al., 2007). There have been conflicting results on the effect of grinding on calibration quality. Some studies found better calibration results for non-ground soils samples (Fystro, 2002; Russell, 2003; Stenberg, 2010). These reports are contrary to those of Barthes et al. (2006) and to our results with a general decreasing calibration error for ground samples. Increasing calibration errors due to grinding had been attributed to the ferocity of the ball mill. This effect had been excluded by using iron-free zirconium oxide grinding jars. The complete destruction and homogenization of aggregates and particles by grinding mostly resulted in lowest calibration errors. Barthes et al. (2006) also found that calibration accuracy (decreasing RMSECV and increasing R^2) was better for ground samples than for sieved ones.

5.1.3 Effect of cup size on NIR calibrations

Throughout different studies the sample presentation for soil NIR spectroscopy is diverse and depends on the spectrometers and their scan units. Some authors took static circular capsules or ring cups (Cozzolino and Moron, 2003; Barthes et al., 2006; Brunet et al., 2007), others

used rotating cups (He et al., 2005) or rectangular sample holders (Chang and Laird, 2002). The dimension of the sample holders is often not even reported. We confirmed with a comparison of two cups with different diameter that cup sizes or rather the scanned area was of minor importance for the calibration accuracy. However, the smallest ring cup we used still had a scanned area of 6.8 cm^2. This agrees with results of Fystro (2002) who did not find significant differences for SOC and N prediction using a smaller spinning ring cup and a larger rectangular transport cup for NIR measurements.

5.1.4 Effect of drying on NIR calibrations

The dominant influence of water on NIR spectra gets reduced when dried soil samples are used. As Wu et al. (2009) mentioned, soil moisture affects NIR spectra appreciably and results for decreasing moisture content of soils lead to increasing reflectance. Whereas drying at a fixed temperature (oven-drying) is the most standardised sample preparation, air drying results in non-steady water content which fluctuates with air humidity. This is especially true for ground soil samples that exhibit a large surface area. However, it is more common for NIR investigations to use air-dried samples (Russell, 2003; He et al., 2005; Brunet et al., 2007) than oven-dried ones (Chang and Laird, 2002; Cozzolino and Moron, 2003). Barthes et al. (2006) investigated the differences of these effects for different grinding levels and discovered no generally valid tendency on calibration quality. Our results showed a better calibration accuracy for oven-dried samples than for air-dried samples for SOC calibration but hardly any instantaneous effect on N calibration (Table 15) which might indicate that functional groups of N are less influenced or masked by water bands. Water plays an even more important role on field-moist samples with higher prediction accuracy for ground, dried than for unground field-moist samples (Malley et al., 2002).

5.1.5 Soil temperature effect

NIR measurements under controlled temperature conditions have rarely been studied for agricultural soils, up until now. In addition, it is not common to mention the temperature range for spectra acquisition. As Kawano and Abe (1995) and Seong et al. (1999) already found out for fruits, sample temperature can have an effect on the performance on NIR. Both studies show that calibrations using samples at constant temperature are not stable in predicting when the sample temperature varied. This is in line with our results for air-dried soil samples, which showed varying calibration accuracy for 24 and 28°C samples, predicted on a 20°C calibration. In addition, our results showed that the effect of temperature on NIR

spectra disappeared with oven-dried samples, indicating that either i) a different soil moisture content caused differences in scanned spectra at different temperatures, or ii) the remaining soil water of air-dried samples was especially temperature sensitive with effects on the whole NIR spectra. Water and hydroxyls have large absorption in the NIR region (Stenberg, 2010) and changing strength of hydrogen bonds definitely influences the NIR absorption (Buening-Pfaue, 2003). According to our results we recommend using oven-dried samples for calibration and prediction since they are relatively resistant towards temperature changes and deliver the best prediction accuracy in comparison to air-dried samples.

5.1.6 Reproducibility of NIR spectra

The precondition for an effective application of NIR for soil chemical analysis is the reproducibility of NIR measurements of soil samples. Once developed, calibrations shall be able to predict additional sample constituents without bias derived from interfering variations in sample conditions (e.g., moisture content, aging), lab conditions (e.g., temperature, humidity) or spectrometer performance. The best correlations towards repeated measures could be achieved for soil samples under sieved air-dried and ground oven-dried condition. These sample treatments are therefore recommended to minimize the time effect on measurements for calibrations. Moreover, our results showed that not standardised dried soil showed a low reproducibility of predicted values out of validation when NIR measurements were repeated. RMSEPs for SOC and N prediction varied independently with time. Not standardised or rather air-dried soil can hold different amounts of moisture. Water and hydroxyls (OH) have strong influences on soil reflectance in the NIR range (Stenberg, 2010) that may lead to worse calibrations results, while organic functional groups are masked by H_2O-bands (Chang et al., 2005; Barthes et al., 2006). This observation is supported by Tekin et al. (2010) who found the soil moisture content to affect the prediction NIR performance for SOC, with lowest calibration error using dried soil samples.

5.2 Field-based NIR measurements

Contrary to pure component systems, soil is a very complex mixture of minerals and organic constituents. The soil properties can vary under different soil mineralogy and their content in soil organic matter. Therefore, international soil spectral libraries, as created by the ISRIC World Soil Information foundation (http://www.isric.org/), make sense in order to gain and collect more information on the variability of soil attributes for different land use, land use changes and climate zones. Generally, minerals and organic matter are the main spectrally

active components of the soil (Ben-Dor et al., 2008). Here, several scientists such as Christy (2003), Viscarra Rossel et al. (2006) and Stenberg (2010) found that chemical and physical soil parameters can successfully be detected via on-line VIS-NIR spectroscopy, which works fast, is easy to handle and delivers satisfactory results. Although, NIR spectra result from overtones and combination bands of primary absorptions in the mid infrared region of the electromagnetic spectrum, the well-established multivariate data analysis helps to find patterns and trends in the NIR as well as VIS regions.

Average VIS-NIR spectra from the three studied field soils showed typical absorption bands (Stenberg, 2010; Tekin et al., 2010) as results from on-line measurements via shank or probe. Differences in spectral baselines and absorption peaks might be based on different measurement directions via tractor-application (horizontal with shank or vertical with probe) and also based on each particular internal structure within the optical unit for spectra acquisition. In addition, changing external influences during the in-field calibration process might have influenced spectra acquisition as well as the error of the systems (shank, probe). Negative influences during scanning might have been thin films of moisture or dust particles on the sapphire window. Even though the manufacturers promise a dust-free driving path through a self-cleaning system, the absence of dust particles cannot be guaranteed 100%. However, before each field calibration, the sapphire window was cleaned with a dry and dust free cloth to minimise interfering influences.

5.2.1 Horizontal field mapping: shank

Overall, the horizontal on-line NIR measurements of the soil worked well for all three fields studied, and R^2 were greater than or equal to 0.9 for SOC and 0.94 for total N calibrations. According to Malley et al. (2002), results with $R^2 > 0.95$ can be rated as excellent, and results with $0.9 \geq R^2 \leq 0.95$ can be rated as successful.

The NIR calibration error (RMSECV) varied between 0.061 and 0.172 for SOC and between 0.011 and 0.016 for N and was lower than in other studies. As one example, Knadel et al. (2011) achieved a RMSECV of 2.37 for SNV-treated spectra and of 6.52 for spectra converted into first Savitzky Golay derivative for a field with a high SOC-gradient. The SOC concentration range comprised 1.44 – 42.9%. Huang et al. (2007) and Munoz and Kravchenko (2011) derived calibration errors for low SOC-gradient fields (similar to the SOC content of the fields investigated in the present study) of 1.89 and 1.65, respectively. In order to achieve satisfactory calibrations with high R^2 and low calibration errors, the amount of calibration samples for large concentration ranges should be accordingly high, and fewer

samples are generally needed for smaller calibration ranges (Naes and Isaksson, 1989). However, the RMSECV and the number of calibration samples are directly related to each other, since the RMSECV calculation formula takes the number of calibration samples directly into account. The more samples used for calibration, the smaller the calibration error. This is possibly the reason why Knadel et al. (2011) obtained high calibration errors in combination with a field study that comprised merely 15 samples for a large SOC concentration range. Presumably, Huang et al. (2007) and Munoz and Kravchenko (2011) have lower RMSECV values than those in the study mentioned since they used more samples for a low-gradient field (Huang et al. (2007) used 79 samples and Munoz and Kravchenko (2011) took 45 samples for SOC calibration).

However, SNV has shown to be the best spectra pretreatment for NIR calibrations in the present study. This follows the results of Knadel et al. (2011), who achieved the highest R^2 and lowest RMSECV for spectra modified with SNV, followed by second derivative, multiple scatter correction, first derivative and raw spectra, in descending order. Moreover, it is common to use SNV as spectra pretreatment (Brunet et al., 2007), since it reduces particle size effects and can remove the linear or curvilinear trends of each spectrum (Barnes et al., 1989).

Next, the calibration results from VIS-NIR spectra were noticeably worse than the results from NIR spectra. Pulling the shank through the irregularities of the soil's surface led to raising and lowering of the optical shank unit, which was placed directly in the topsoil. The optic was briefly lifted out of the soil several times per measurement campaign, when it would detect the visible spectrum of the sunlight. On the basis of NIR spectra and calibrations, the VIS region was not considered, and on the basis of VIS-NIR spectra and calibrations, the VIS region was involved in calculations. Depending on the default settings and the sensitivity toward varying proportions of visible light, which is a black box for all VERIS spectrometer software user, the average spectra (for 3 m of travel) were automatically detected as outliers or not. When they were left in the spectra plot for further processing, worse calibration results might have been the consequence.

However, the VERIS spectrometer software chose 'good' spectra by separating out those with a Mahalanobis distance greater than three. This was the default value of the system and therefore not changeable. However, this setting is common when outliers are to be identified (Shenk and Westerhaus, 1991; Barthes et al., 2006; Brunet et al., 2007; Knadel et al., 2011). The possibility to extract 'bad' soil/light spectra would require a change in the Mahalanobis

distance settings: The value should be reset to a lower value. But this is not possible at present.

5.2.2 Vertical field mapping: probe

On the whole, probe examinations via VIS-NIRS showed successful to excellent calibration results for FALNO (SOC: $0.9 \geq R^2 \leq 0.98$; N: $0.91 \geq R^2 \leq 0.98$) and HdBA (SOC: $0.91 \geq R^2 \leq 0.99$; N: $0.91 \geq R^2 \leq 0.99$), according to Malley et al. (2002). Studies on Espenberg resulted in worse calibration results, with partly $R^2 < 0.5$ and were not satisfactory. In addition, the RMSECV values for SOC and N and for each depth section were almost always larger for Espenberg, when compared to calibration results from FALNO and HdBA. This can be attributed to a lower data density per area unit and depth section for Espenberg (9 field spectra ha^{-1}) than for the other fields (FALNO: 64 field spectra ha^{-1}, HdBA: 24 field spectra ha^{-1}). Reasons for Espenberg´s lower data density were probably the high average tractor speed (Table 4) and also the automated field spectra removal of 54.28% performed by the VERIS Spectrometer software (Table 18). A reason for that field spectra removal might have been clayey parts of the field soil that hampered the self-cleaning of the optical unit during measurement.

Table 30: Calibration results of SOC obtained from four field studies that performed on-line NIR or VIS-NIR measurements.

Author	Field size	SOC concentration range [%]	R^2	RMSECV	Measurement range
Munoz and Kravchenko (2011)	12 ha	0.69 - 1.40	0.5	0.154	NIR
Nocita et al. (2011)	130 km transect	0.18 - 6.03	0.83	0.525	NIR
Nocita et al. (2011)	130 km transect	0.18 - 6.03	0.83	0.526	VIS-NIR
Knadel et al. (2011)	12 ha	1.17 - 38.32	0.97	2.37	VIS-NIR

However, no review has been conducted that classifies successful calibrations according to the size of the calibration error (RMSEVC). Therefore, a comparison of the present RMSECV values with those from the literature was advisable. Table 30 shows the R^2 and RMSECV values from four field studies developing SOC calibrations from on-line measurements. Overall, the RMSECV values ranged between 0.154 and 2.37 and the R^2 between 0.5 and 0.97.

The RMSECV values from the single-depth calibrations of FALNO and HdBA were lower than or equal to the lowest RMSECV value of 0.154 in the studies mentioned (Table 30). However, Munoz and Kravchenko (2011) attained a similar calibration error of RMSECV = 0.154 for a calibration with a low SOC content, such as in the present study. Nocita et al. (2011) and Knadel et al. (2011) obtained larger RMSECV values than those of FALNO and HdBA. Nocita studied a 130 km transect comprising six vegetation types with different rainfall ranges. As is well-known, soil moisture considerably affects the calibration and prediction performance of SOC. When compared to dry samples, wet samples delivered worse calibration results (Tekin et al., 2010) so that R^2 of ≤ 0.89 and RMSECV of ≥ 0.67 were common calibration results for a soil with 5–30% gravimetric moisture content (Rodionov et al., 2015). Additionally, Tekin et al. (2014) performed on-line NIR measurements on a field with different vegetation index data and attained moderate calibration accuracy of SOC with R^2 of 0.75 and RMSECV of 0.17. So both variables, the varying moisture content and the vegetation index data, might have led to worse calibration results of SOC in the study by Nocita et al. (2011). The large RMSECV obtained by Knadel et al. (2011) possibly originated from a field study with too few reference samples for a soil with a high SOC gradient. As Davies and Fearn (2006) and Bellon-Maurel et al. (2010) found, RMSECV is dependent on the measurement range and the amount of reference samples for calibration. However, more work has to be done in order to classify calibrations as successful, especially with respect to RMSECV. A measurement protocol that standardises field measurements via on-line NIRS would be helpful as a uniform measurement background for calibration.

5.2.3 Data interpolation

Kriging proved to be a good method for extrapolating a large number of SOC stock data calculated out of results from on-line NIR measurements. In general, Kriging can interpolate well between stationary environmental elements (Li et al., 2013), and it works by incorporating the spatial autocorrelation in the regression residuals into the Kriging system (Hengl et al., 2004). Since Kriging is basically an interpolation technique, it tends to smooth the variable of interest (Mishra et al., 2012). To perform successful linear regressions for Kriging, a sufficient number of field data is necessary. Interpolation results based on 12 SOC concentrations, as a reference for on-line probe NIR measurements and calibrations, showed that Kriging was not the method to be applied. The model data were probably from too few soil samples. In line with Mishra et al. (2012), our 12 and also their 18 reference SOC

concentrations were not enough soil data to create a semivariogram. Moreover, the predicted field data that was based on probe measurements were too few to create a semivariogram. Even though a plurality of field spectra were scanned during measurement, for each depth section 36 field spectra were scanned on Espenberg, 68 on HdBA and 142 on FALNO. These results mainly agreed with those of Webster and Oliver (1992), who advised the use of more than 100 data points for successful Kriging – having fewer data points than 100 gave unreliable results, or semivariograms could not even be created. In a recent study by Oliver and Webster (2014), they suggested a minimum of 100–150 data points for Kriging, depending on soil variability. Keeping in mind that the on-line NIR spectroscopy is used as a time-effective *in situ* soil measurement, a larger number of probe measurements that was carried out in the present study would waste measuring effort and would not be time-effective. Since Kriging could not be applied for our probe data, it was put aside as a method for comparison between shank and probe analyses. However, the results from shank analysis alone could be evaluated well by Kriging.

5.2.4 Calculation of SOC stocks and their errors

SOC stocks calculated out of soil analysis without on-line NIR measurements varied from 38.0 (smallest HdBA SOC stock) to 74.1 Mg ha^{-1} 30 cm^{-1} (largest Espenberg SOC stock). In comparison to Nyssen et al. (2008), who quantified SOC stocks for a cultivated land in the Ethiopian rift valley with 33 Mg ha^{-1} 20 cm^{-1}, Sleutel et al. (2006), who studied SOC stocks for a Flemish cropland in Belgium with 50 Mg ha^{-1} 30 cm^{-1}, and Leifeld et al. (2005), who calculated SOC stocks for arable land in Switzerland with 50.7 Mg ha^{-1} 20 cm^{-1}, HdBA SOC stocks ranged in the middle of mentioned SOC stock data. FALNO, and Espenberg held larger amounts of SOC stocks in comparison to HdBA. This is probably due to a higher mean SOC concentration in the 0–30 cm layer of FALNO with 1.32% SOC and with 1.4% C for Espenberg in comparison to the mean SOC concentration of HdBA that amounted to 0.78% SOC. This is in line with Schrumpf et al. (2011), who estimated SOC stocks with a magnitude of 86.5 Mg ha^{-1} for Gebesee, Germany, 82.4 Mg ha^{-1} for Grignon, France and 60.4 Mg ha^{-1} for a field in Carlow, Ireland, with generally higher mean SOC concentrations in the upper 30 cm (all have < 5% SOC on average in the upper 30 cm soil layer).

Overall, SOC stocks in the present study varied from field to field as well as between different models for SOC stock assessment per field, partly due to the different calculations for BD assessment for directly measured depths and for SOC data acquisition. As is well known, SOC stock assessments are generally associated with large uncertainties. These may

impair the detection of SOC stocks as well as the identification of the main driving forces (Falloon et al., 2006; Ogle et al., 2006). Furthermore, the influencing factors of SOC in different geographical locations are difficult to determine, resulting in complex and also uncertain spatial distribution characteristics of SOC (Kumar and La, 2011). Uncertainties are difficult to identify and to quantify because they come from complex interactions between the variables involved in SOC stock assessment (SOC concentration, bulk density, sampling depth and rock fragment content). The large relative variability of SOC concentration contributes to 84–99% to SOC stock uncertainty, whereas BD variation only contributes with < 5% to SOC stock uncertainty (Holmes et al., 2011).

Nevertheless, BD is conventionally difficult and time-consuming to measure, and sample collection can be complicated by dry soil, coarse textured soil or coarse fragments in the soil. Besides the conventional Kopercki ring method, BD can also be estimated through the use of pedotransfer functions or via on-line measurements. The most accurate results can be received by obtaining data via the Kopercki ring method in the laboratory. However, capturing accurate field variability can result in a large number of ring samples with a subsequent time-consuming and costly laboratory analysis (Andrade-Sanchez et al., 2007). Therefore, the ring method is inefficient if a dense coverage of data points is required for a complete assessment of a field's bulk density (Donohue et al., 2013). Therefore, the use of pedotransfer functions simplifies SOC stock estimation since the laboratory element of BD assessment is then phased out.

When Callesen's pedotransfer regressions (Callesen et al., 2003) were used for BD estimation, the resulting SOC stocks (out of direct chemical analysis of SOC, without NIR) had the highest errors (SE_{REF} and SE_{MEAS}) in comparison to all other SOC stock calculation models. Even having selected the most suitable BD regressions out of six possibilities (according to soil texture), the high errors delivered questionable results. As Boucneau et al. (1998) and De Vos et al. (2005) found that indirect BD estimates based on pedotransfer functions can lead to errors from 9% up to 36% of the SOC stock. In addition, Schrumpf et al. (2011) found that using PTFs for estimating BD led to inaccurate or biased SOC stocks and that direct measurement of BD was laborious but led to more accurate results. Apparently, BD functions should be used with great caution, and more work is needed to clarify the various impacts on BD estimations (Hollis et al., 2012). Estimating BD via a cylinder auger is in common use but the device is also not perfect. Pushing the augers into the soil usually leads to soil disturbances at each depth and the average value of bulk density is being reduced. Pushed core samplers, as used in the present study, can decrease the average value of bulk density by

0.04 Mg m^{-3} (Raper and Erbach, 1987). Therefore, compaction correction was performed for the compacted area, which could be visually detected at a depth of 40-60 cm. On-line measurements for BD prediction can be realised by using the on-line sensor for topsoil BD prediction used by Quraishi and Mouazen (2013). The sensor has been available since 2013 and has shown satisfactory results with $R^2 = 0.6$, when maps of topsoil BD measured from core samples were compared with maps of BD predicted using the on-line sensor. VIS-NIR spectra were used to predict soil BD. Since VIS-NIRS can only detect chemical structures, an algorithm was developed that delivered correction factors for a BD prediction model, including moisture content and soil particle size fractions. The on-line BD sensor was shown to be capable of rapidly predicting field dry BD for a large number of samples, compared with traditional methods.

SOC stocks predicted on the basis of on-line NIR measurements delivered an overall SOC stock range of 41.7–71.4 Mg ha^{-1} 30 cm^{-1} and were smaller than estimated on the basis of direct chemical analysis (38.0–74.1 Mg ha^{-1} 30 cm^{-1}) in the present study. Also, the range of the predicted SOC values, needed for SOC stock calculation, was mostly smaller than the range of the reference values (Figure 16). Generally, data can be reliably predicted for a SOC range that has been calibrated previously. Consequently, FALNOs and HdBAs SOC data that were predicted outside the calibration range must be questioned and require careful interpretation. Especially for FALNO, out of 4807 predicted SOC concentration values, 291 were predicted below the reference concentration range. The difference to the lowest calibrated concentration value was 0.08% (minimum SOC value out of calibration: 0.99%, minimum SOC value out of prediction: 0.91%). Possibly, those low values were not captured via punctual reference soil sampling but via area-wide on-line NIR measurements scanning the soil and delivering a plurality of field spectra: with 4771 more measurements via shank application and 2982 more measurements via probe application, when compared to the number of collected reference samples. The difference of the data ranges between reference and predicted SOC concentrations and SOC stocks for HdBA and Espenberg is lower, indicating a satisfactory reference sampling for both fields.

5.2.5 Time aspect of SOC stock estimations

Neither type of on-line NIR measurements - those carried out via shank or by probe - differed considerably in regard to working time and costs for SOC stock assessment. Nevertheless, the results of the shank were obtained in a shorter time with lower costs, revealing a two-dimensional view of the soil. Via probe, a three-dimensional view into the soil was obtained.

Adding the third dimension required more time and higher costs, but it delivered more information about the soil. Since SOC is not predominantly stored in the topsoil, but large amounts may be stored in subsoil horizons below 30 cm depth (Lorenz and Lal, 2005), SOC stock assessments via NIRS with a probe present a very meaningful option to respond to the Kyoto Protocol's second commitment period 2013–2020.

In an overall comparison of SOC stocks, including their errors and the required time and costs, no clear recommendations can be given as to whether to use conventional soil sampling or modern on-line soil measurements via shank or probe for SOC stock assessment. Each method has its advantages and disadvantages. Nevertheless, soil science is developing in the direction of replacing traditional laboratory methods with fast, cost-effective and easy-to-handle recording technologies (Munoz and Kravchenko, 2011). NIR is among the most popular tool (Malley et al., 2002) due to its strong potential to detect soil carbon (Viscarra Rossel et al., 2006). Furthermore, numerous high density data can now be obtained at a relatively low cost (McBratney et al., 2006). Investigations of both implementations should take place on more fields with a larger variation in texture and organic carbon content to better determine the power of shank and probe recordings for SOC stock assessment.

5.3 Field reproducibility study

5.3.1 Soil NIR spectra as well as calibration results for SOC

The field soils under study were measured twice via on-line NIR spectroscopy in order to determine the reproducibility of the measurement method. The first results were a plurality of soil spectra per field. In combination with the SOC content measured via direct chemical analysis, calibration models were performed, delivering calibration accuracy data, such as R^2 and RMSECV, as well as calibration model quality data, such as RPD, RER and SEC. The calibration results were considered to be good, according to Chang et al. (2001) and Couteau and Schaller (2003) with regard to $R^2 > 0.8$. In the successful analysis of agricultural commodities, it is even desirable to have $R^2 > 0.95$ (Tekin et al., 2010), but samples of complex material, such as soil with variable composition, can have lower R^2 with satisfactory results as well (Viscarra Rossel et al., 2006). In a comparison between the calibration results of the present study and those from other studies and similar data acquisition, Knadel et al. (2011) achieved a lower R^2 of 0.76 and a larger RMSECV of 6.59. RMSECV values in the present study ranged between 0.39 and 1.31. Also, Mouazen et al. (2007b) reached lower calibration accuracy, with an R^2 of 0.74 and an RMSECV of 4.8. All mentioned studies worked with soil data comprising different SOC ranges: Knadel et al. (2011) studied one field

with 11.7 to 383.2 g kg^{-1} SOC and Mouazen et al. (2007b) several fields with a total range of 7 to 60 g kg^{-1} SOC. Since the R^2 and also the RMSECV are dependent on the measurement range (Davies and Fearn, 2006; Bellon-Maurel and McBratney, 2011), these values were not the best choice for measuring the merit of a calibration. However, these values are very popular for evaluating NIR calibrations throughout the literature. In addition, the RPD and the RER values are also widely adopted by the soil community, and larger values of RPD and RER indicate better fits of calibration models. However, they also depend on the measurement range and the distribution of the SOC data (Bellon-Maurel et al., 2010). Therefore, the large SOC measurement ranges from Knadel et al. (2011) and Mouazen et al. (2007b) led to larger RPD values and to larger RER values (Knadel et al., 2011: RER = 5.64, RPD = 2.11, Mouazen et al., 2007: RER = 11.04, RPD = 1.97) than the RPDs and RERs calculated in the present study.

Nevertheless, Malley et al. (2002) stated that RPD is the statistically least sensitive measure when a concentration range is increased by only a few values. It is therefore suitable for rating the present calibration results having data with negligible SOC changes for repeated measurement campaigns. The SOC calibrations of HdBA and FAL250 delivered similar R^2 for repeated measurement campaigns (measurements, -1- and -2-, Table 25) whereas the RMSECV, SEC, RPD and RER values varied for repeated measurements. As can be seen in Table 10, the measurement ranges of SOC differed from the first measurement to the second measurement for each field. This obviously led to different calibration accuracies and calibration model quality data since these values are generally dependent on the measurement range. However, even the measurements on FAL250 with the same SOC-data range (-1- and -2-^{rep}) did not lead to similar calibration results. Possibly, the NIR spectra acquisition might have been influenced by varying soil moisture and soil temperature (Mouazen et al., 2007b; Tekin et al., 2010; Kuang and Mouazen, 2011), leading to different NIR spectra per measurement campaign. In addition, different measuring depths (7 cm +/- 3 cm), resulting from encountering stones and old ploughing materials (two times), might have led to NIR spectra including different soil information. The possible consequence might have been different PLSR components for SOC calibrations of repeated measurements (different plethora of spectral information), which is shown in Table 25. However, the chosen number of factors described most of the variation in the spectra. Since all remaining factors resemble noise, they are usually ignored (Viscarra Rossel et al., 2006).

Up to now, there has been no comprehensive field study taking into account all the influences for SOC calibration at once. A measurement protocol should be established that ensures a

successful SOC calibration by standardising the steps of the calibration process, soil sampling strategy and in-field spectra acquisition. For this purpose, a field study should be repeated at least 10 times so that random effects and significant influences on SOC calibration can be better estimated.

5.3.2 SOC concentrations and SOC stocks: Calculations on the basis of direct chemical analysis versus on-line NIR measurements

The application of a mobile field spectrometer, performing on-line NIR measurements, was used to predict average SOC concentrations and average SOC stocks at the field scale. A reference SOC dataset served as a calculation basis for the SOC stocks (TEST) predicted out of the NIR spectra scanned on-line and for reference SOC stocks (REF) estimated without the use of the mobile field spectrometer for comparison purposes. The average reference SOC dataset from FAL250 and HdBA (FAL250 ~ 15 g kg^{-1} C, HdBA ~ 8.1 and 9.3 g kg^{-1} C) is typical for German agricultural soils that have till as geological parent material (Wessolek et al., 2008). Due to a large variation of SOC contents, the mentioned study generally distinguished between low, medium and large average SOC concentrations. The fields studied (HdBA and FAL250) belong to their determined medium concentration range of 7 to 16 g kg^{-1} C for soils with subglacial till as parent material. In a comparison between the results from direct chemical analysis and those received through on-line NIR measurements, the predicted average SOC concentrations and SOC stocks were consistently larger than the corresponding data of the reference analysis. This is in line with Mouazen et al. (2007b), who obtained a lower average concentration for total carbon, with 0.99 g kg^{-1} C out of the collected soil samples measured in a laboratory and a higher average total carbon concentration of 1.05 g kg^{-1} when the measurements were carried out on-line (VIS-NIR). This can be attributed to the higher variation and larger range of the total carbon recorded via the on-line measurement.

The standard errors of SOC concentrations and SOC stocks were consistently smaller when generated out of on-line measurements and larger when computed out of direct chemical analyses. This was due to the small number of soil samples used for reference analysis in contrast to the large quantity of spectra scanned during the NIR measurements, both influencing the calculation of the error sizes. The coefficients of variation of SOC and its stocks were larger for the predicted than for the reference values with respect to FAL250. This can be attributed to larger standard deviations of the wider SOC and SOC stock ranges of the predicted values. For HdBA, the reference SOC and SOC stocks had larger CVs than the

corresponding predicted values. Here, the smaller predicted data range is responsible for the lower CVs and the larger reference data range for the higher CV´s. With regard to the repeated measurement campaigns on FAL250, the difference of SOC and SOC stock variation was little or even none (Table 26). Considering HdBA, the variation of SOC and SOC stocks was clearly different for the repeated measurements in five of six cases. For all SOC and SOC stock computations, the reference SOC dataset served as the basis for calculation. A closer look at the reference dataset revealed different soil sampling strategies per field – the nested design for FAL250 and the cluster design for HdBA. The reference sampling on FAL250 could be realized by using similar field positions for the sampling during both measurement campaigns. The cluster design carried out on HdBA did not lead to similar field positions within repeated measurement campaigns. Here, representative field positions were chosen according to the spectral feature scanned per measurement. Here, the spectral information received via the on-line measurements was clearly different for repeated measurement campaigns, which likewise influenced the SOC calibrations and the size of SOC stocks, leading to different variation coefficients of SOC concentrations and SOC stocks.

However, generally, widely used sampling methods include grid, random and cluster sampling as well as the nested design, but results for a 'best-grid-size' for various soil components at the field scale still remain inconsistent. Even a comparison between different sampling methods did not show uniform results. Anami et al. (2008) state that nested sampling is more suitable for a range determination than grid sampling for areas where the space variability of properties is not known and needs to be studied (field scale). Hedley et al. (2012) found that random sampling is a suitable method for soil carbon monitoring at the national scale, since reference analysis and modelled values show no significant differences. From the present study, no conclusions can be drawn in regard to the 'best sampling method' for obtaining successful SOC calibrations and accurate SOC stock assessments.

5.3.3 Semivariogram, Kriging and MDD results

Kriging, as an advanced geostatistical procedure that generates an estimated surface from a scattered set of points, assumes that the spatial variation is statistically homogenous throughout the surface. The spatial variation is quantified by the semivariogram, which is one of the most suitable statistical techniques to signify spatial correlation (Cahn et al., 1994; Gupta et al., 1997). It is applied to quantify the degree of similarity of a pair of samples at different locations, separated by a certain distance (Kerry and Oliver, 2004). In general, a closer distance suggests a stronger similarity of the samples. The spatial correlation

disappears when the interval of two samples is more than the range. In this case, samples cannot be used for temporal or spatial interpolation (Wang et al., 1999; Kerry and Oliver, 2004; Javed et al., 2005). Overall, ranges varied between 36 and 74 m for SOC concentrations and between 38 and 91 m for SOC stocks for the studied fields and all calculation variants. There are many studies that have worked on the spatial dependency of SOC (Rueth and Lennartz, 2008; Santra et al., 2008; Tornquist et al., 2009; Worsham et al., 2010; Meersmans et al., 2011; Liu et al., 2012; Mondini et al., 2012; Ogunwole et al., 2014; Kumar, 2015), but the SOC ranges were different for all examined sites and sampling methods.

For further semivariogram analysis, a ratio between nugget and sill was recommended by Cambardella et al. (1994) and Javed et al. (2005) for use as a measure of spatial dependence and for small-scale and sampling-induced variability: Here, a ratio of <25% indicated strong spatial correlation, a ratio of 25% to 75% signified moderate spatial correlation and a ratio >75% indicated weak spatial correlation. The ratio of nugget to sill ranged from 46% (FAL250) to 71% (HdBA), indicating that the spatial correlation of SOC concentration was moderately dependent. Normally, a low nugget to sill ratio indicates a small degree of small-scale variation and high spatial autocorrelation. In contrast, high nugget to sill ratios of ≥84% pointed to a weak spatial correlation of SOC stock FIELDSPEC for FAL250. Low nugget to sill ratios for HdBA (first measurement campaign) pointed to a strong spatial correlation of both SOC stock types. Kriging generally provided estimates at unsampled (or unscanned) sites: Its interpolation provides a best linear unbiased estimation for quantities that vary in space (Oliver and Webster, 2014). We used punctual Kriging with a resolution of 4 x 4 m. The results for the repeated measurements on the two fields were four Kriging maps showing the distribution of SOC concentration (Figures 22, 23).

The results for FAL250 were two Kriging maps (first and second measurement campaigns) with a similar distribution of SOC content. Apparently, the measurement campaigns, repeated within a short time distance of two weeks, revealed comparable Kriging results of SOC distribution. This also suggested a good reproducibility when on-line NIR measurements were repeated. The Kriging results for HdBA were two maps with different distributions of SOC content. Areas with small and high SOC concentrations were received at equal locations within repeated measurement campaigns, but SOC variability in intermediate areas varied between measurement to measurement campaigns. Even though the 4 x 4 m grid used led to the same amount of grid points for Kriging map generation, the amount of predicted SOC data, integrated into the Kriging calculation, differed remarkably – between 6288 (number of cleaned field spectra, first measurement campaign) and 1582 (number of cleaned field spectra,

second measurement campaign) SOC values. The most important reason for the low number of SOC values received from a low number of cleaned field spectra from measurement two was probably due to the soil tillage carried out before the measurement started. The farmer loosened the soil via the use of a grubber some hours before the beginning of measurement two. Later, during the measurement, the NIR spectrometer automatically stopped the measurement several times, due to vibrations caused by a rough soil surface. The consequence was a low number of field spectra scanned during that measurement. When the measurements were repeated on the same day for comparison, the number of scanned field spectra and cleaned field spectra was always that low. Despite soil tillage, the difference map of Kriging of HdBA showed a maximum of SOC concentration difference of 0.8 g kg^{-1}, which indicated only small changes of SOC from the first to second measurement campaign. Since no significant changes in SOC were assumed for a time period of two years, the result was compliant with the initial assumption.

MDD was used as a powerful tool to determine the quality between different measurement methods (on-line NIRS vs. direct chemical analysis) and repeated measurement campaigns. Generally, MDD estimations revealed detectable changes of the SOC concentrations between 0.2 and 2.1 g kg^{-1} C as well as changes of C stocks between 1.3 and 8.1 Mg ha^{-1} 30 cm^{-1}. Generated out of on-line measurements, the MDD values of SOC concentrations and SOC stocks were consistently lower than those generated out of direct chemical analysis. That was mainly due to the standard error that was incorporated into the MDD assessment. The high number of measuring points attained via the on-line measurement led to small SEs and consequently to low MDD values. Accordingly, the low number of soil samples – collected for direct chemical analysis – led to higher SEs and to higher MDD values. With regard to the repeated measurement campaigns and the number of soil samples or field spectra for MDD assessment, an equal amount of soil samples was only obtained by the conventional sampling followed by laboratory analysis. Using on-line NIRS, a different number of field spectra was gained from each measurement campaign, which was incorporated into SOC concentration, SOC stock and MDD assessments. Therefore, and at first glance, the conventional method seemed to be more suitable for verifying the quality of the repeated measurements. But here, only a few samples were taken as representative for the field under study. A low sampling intensity can result in Type II statistical errors so that a difference at a given significance level will not be determined when there really is one (VandenBygaart et al., 2007).

When Garten and Wullschleger (1999) examined the MDD of SOC for switchgrass plots in the United States, they found out that a large sample size of more than 100 samples is needed

to detect SOC stock changes of 2% to 3% with a good statistical power of 90%. Of course, this sample size is a challenge for monitoring SOC stock changes. Therefore, and due to SOC variation, the on-line method is a promising tool for gaining more soil information with a plurality of field spectra per field than via conventionally sampled and analysed soil. Indeed, the difference in MDDs from the repeated measurement campaigns was lower for the on-line method than for the conventional sampling method. That was an asset for the on-line NIRS. However, the conventional and on-line methods both include analytical and methodological errors that influence measurement results to different degrees. Particular errors were not determined in this study, so the present results were not absolute values, but rather first indications for a comparison between repeated measurement campaigns of SOC concentrations and SOC stocks. It is not easy to fully estimate all errors (Lischer, 1993). Sources of errors could have been different soil conditions during measurement, such as soil moisture, temperature, compaction, surface roughness and root penetration, as well as different soil textures, different soil sampling methods for SOC assessment, a different calibration of spectrometers for laboratory and in-field measurements, different numbers of soil samples or field spectra and many more (Naes and Isaksson, 1989; Chang et al., 2001; Wu et al., 2009; Stenberg, 2010; Kuang et al., 2012; Rodionov et al., 2015). It is advisable to perform a field study with at least 10 replications of measurements so that important analytical and measurement errors can be quantified and be taken into account in measurement results and following calculations.

6 Conclusions and outlook

The results presented and discussed in the previous sections demonstrated the great potential of NIRS for laboratory and field soil analyses. This section introduces the conclusions derived from the results as well as the outlook on desired further developments in the area of soil spectroscopy.

6.1 Laboratory-based NIR measurements

The application of NIR spectroscopy in the laboratory to estimate SOC and N content operates at the technical limits of the method due to low concentrations of the target components and a very high variability of the sample matrix mainly comprising of different minerals and oxides in different size classes. This study was carried out in a broad range of soil types from Northern Germany, investigating several possible parameters during sample preparation and measurement that influence the calibration accuracy. However, parameters such as number of replicate spectra or size of the measurement cup were found to have only minor impacts on calibration results. Best calibration results for SOC according to sample treatments could be achieved by using ground and oven-dried samples measured in a big cup. Whereas N was not very sensitive to different forms of drying, calibration results for SOC were significantly better for oven-dried samples. Additionally there was a positive effect of grinding, for N and for SOC ground samples even led to significantly better results. Calibration of soil components seemed to vary in sensitivity to sample preparation. Different grinding levels were found to have the largest influence on calibration results. Since not standardised air-dried soil samples showed a low reproducibility within nine months, standardised oven drying of soil samples is strongly recommended for NIR measurements. Furthermore, oven-dried samples were less sensitive to variations in the laboratory temperature.

6.2 Field-based NIR measurements

While the spatial and temporal variability of SOC and total N makes accurate quantification more difficult, the fast NIRS was tested on-line for its feasibility to estimate SOC concentrations and stocks.

In order to obtain successful calibration results for SOC and N via the VERIS shank it is worth using the NIR region of field spectra and SNV as spectra pretreatment. It is not

advisable to use VIS-NIR spectra since they led to worse calibrations, even with SNV-treated spectra.

Receiving successful to excellent calibration results via the VERIS probe, the VIS-NIR region should be used as well as reflectance spectra or those pretreated by the first derivative.

For universal calibrations including the whole scanned soil depth (0–70 cm), SNV-treated spectra led to the best calibration results for SOC and N.

No recommendations can be given for a best SOC stock model for SOC stock estimation and the stocks varied from shank to probe application, between different BD estimations and different data usage for the soil depths.

However, using the VERIS shank for horizontal data acquisition, the errors of SOC stock models were consistently lower (SE_{PRED} and SE_{RMSEP}) or predominantly lower (SE_{MC}) than those estimated via probe investigations. Furthermore, slightly less working time and costs were needed when shank-based SOC stock estimations were on task, and the time difference compared to probe-based SOC stock assessments was small.

6.3 Field reproducibility study

Due to variations in SOC at the field scale, it is a challenge to determine SOC stocks and SOC stock changes in a rapid and cost-effective way. Before changes of SOC content can be clearly detected, it is essential to examine the reproducibility of the measurement method, here the predictive accuracy of repeatedly carried out on-line NIRS investigations.

For estimating SOC calibrations from repeated measurements with satisfying results it is worth using the unmodified reflectance spectra scanned via on-line NIRS. Calibration accuracy and calibration model quality data such as RMSECV, RER and SEC were not suitable for rating SOC calibrations from repeated measurement campaigns since they are generally sensitive towards small differences in the reference SOC concentration ranges. In almost all cases, it was more profitable to establish a field-specific regression for BD and SOC stock assessments than to use a BD regression out of the literature.

Since a higher variation and a larger range of SOC can be recorded via the on-line measurement, predicted SOC concentrations and SOC stock data were primarily larger than those computed out of conventionally sampled and analysed soil samples. It was not advisable to rate repeated measurements via the standard error of SOC, since it was sensitive to the amount of field data.

It pays off to use NIRS on-line, because the variation of predicted SOC concentrations and stocks was similar for the most part and detectable differences of SOC concentrations and

stocks were smaller, and the advisable time interval between repeated measurement campaigns was calculated to be shorter than results revealed from the conventionally direct chemical analysis. Furthermore it is advisable to collect and analyse soil samples for each measurement campaign, since a repeated usage of SOC data led to worse SOC calibration results, even though relative MDDs were similar or differed slightly between all measurement campaigns and types. The calculated MDDs should be revised – adding more important error data for MDD assessment, such as soil sampling errors and SOC analysis and NIR measurement errors, to increase the accuracy of MDDs.

6.4 Comparison between laboratory and field instruments and measurements

The reduction of analytical costs when NIRS is used has been reported to be 80% (Foley et al., 1998) or even 80–90% (Madari et al., 2005). This is a major reason why there is increasing interest in minimising conventional laboratory analyses and upgrading cheaper proximal soil sensing methods for field use, in particular, with reflectance spectroscopy. When SOC and other soil components are measured spectrally, the accuracy level differs depending on the type of spectroscope constructed for laboratory or field use (Stevens et al., 2008). A decrease in accuracy was discovered by some authors when laboratory measurements were replaced by field measurements but, however, the performance of on-line sensors is regarded as sufficient for most precision farming applications (Gubler, 2011). Generally, the performance of NIR measurements depends on the type of instrument, its resolution, the environmental conditions and many more factors (Stevens et al., 2008).

There are diverse NIR instruments for laboratory and field use classified into several types, depending on different light-splitting principles. Examples of NIR instrument types include scanning grating, FT, filter and acousto-optical tunable filter, of which FT and grating spectrometer are the widest applications. The latter were also used in this study: an FT spectrometer for laboratory measurements and a grating spectrometer for on-line field use. An FT spectrometer can deliver preferable wavelength accuracy and spectral resolution within a limited wavelength range, since it measures one wavelength at a time and scans throughout the range of wavelengths to complete a spectrum. Grating spectrometers are the best choice for quantitative measurements across a broad wavelength range with high signal to noise ratios because they capture the whole spectrum simultaneously. However, FT spectrometers are very sensitive to vibrations since they operate with moving mirrors, based on the principle of the Michelson interferometer, and are therefore unsuitable for on-line field use (such as the

MPA from Bruker, used in this work). Moreover, some NIR spectrometers designed for the laboratory can be transported to a site even though they are not designed for this (unfavourable for outside use: moving mirrors, the need for CO_2–free air or dry air) (Wang et al., 2015). Besides the mobile variant of NIR spectrometers that measure static and *in-situ*, such as the portable ASD Fieldspec spectroradiometer (Analytical Spectral Devices, Boulder, Colorado, USA) and the FOSS NIR System spectrophotometer (FOSS NIR Systems, Silver Spring, Maryland, USA), the on-line variant delivers much more field data within a short amount of time, which is needed for precision agriculture. Only three NIR sensors for intensive soil mapping are available today that measure on-line and *in-situ* (one was used in this study – the VERIS system), which can probably be attributed to difficulties in designing a sensor that can penetrate the soil and acquire spectral data successfully (Mouazen et al., 2007a). Even though typical field conditions, such as varying soil moisture content, soil structure, stoniness and also coarse organic residues are less desirable, there are promising results using on-line NIRS for the quantitative determination of SOC and total N content (Ben-Dor et al., 2008; Viscarra Rossel et al., 2009; Knadel et al., 2011).

All field-based instruments are grating spectrometers with an indium gallium arsenide short wave infrared system (such as FOSS NIR Systems and VERIS Technologies Inc.) or a lead-sulfide (PbS) detector system (such as ASD Fieldspec) and a fixed resolution of 1.4–2 nm (ASD Fieldspec), 2 nm (FOSS NIR Systems) or 6 nm (VERIS Technologies Inc.). The laboratory-based FT instrument has a single PbS detector with variable resolution, which was set at 1 nm for spectra acquisition. In a comparison between laboratory- and field-based quantitative measurements, the difference in the spectral resolution may be of less importance. As Armstrong et al. (2006) found, the results of a quantitative analysis of several properties in wheat indicated comparable performance for a 10 nm grating instrument and an 8 cm^{-1} FT instrument. Additionally, in a comparison between quantitative measurements with an FT MPA-based system or a pre-dispersive grating system, McArthur and Greensill (2007) determined that both systems deliver similar predictive accuracy for the kaolinite content in Weipa bauxites. Likewise, Calderon et al. (2007) concluded that FT NIR instruments are comparable to grating in their potential to build calibrations for fatty acids. However, more efforts have been made to compare between the laboratory- and field-based measurements of soil properties, but calibration and prediction results are conflicting. Stevens et al. (2008) stated that the performance of field spectroscopy is equivalent to laboratory spectroscopy when measuring SOC under specific surface conditions, such as the absence of vegetation and a low variation in moisture content. They performed VIS-NIR measurements with the same

ASD Fieldspec spectroradiometer for laboratory or field use. Further studies revealed a decrease in accuracy passing from laboratory to field conditions. Oltra-Carrio et al. (2015) investigated the improvement of soil moisture retrieval from VIS-NIR data and found good statistical performance for the calibration of soil moisture content with laboratory spectra. The subsequent validation performed better for laboratory spectra than for in-situ spectra in 60% of the cases. The authors considered five different approaches of signal processing and four criteria, named spectral indices. Nocita et al. (2011) observed small differences in the mean reflectance values of laboratory and field VIS-NIR spectra and found a consistently better calibration and prediction accuracy (larger R^2 and lower RMSECV/ RMSEP values) of SOC – for the first 5 mm of the topsoil – for laboratory than for field calibrations.

In contrast, results from the present study reveal a better calibration performance from field spectra than from laboratory spectra: The RMSECVs are smaller for the field calibrations in 28 of 31 cases. This might be attributed to three factors affecting the calibration to different degrees: the moisture effect, the variability of the reference data and the instrument calibration. Even though there was no precipitation prior the field measurements (within one week before measurement), the water absorption peak at around 1900 nm shows the presence of soil moisture (see figures 15 and 20). Due to moisture, there was certainly an information loss regrading the organics. As Mouazen et al. (2006) found, the NH and CH peaks of organic functional groups - present in their field spectra scanned on-line – were masked by water absorption. The non-visible absorption of the organics at around 1900 nm might have led to smaller SOC calibration errors in this study, since less spectral data had to be calibrated, on the premise that the spectral changes of soil moisture were low. Moreover, the number of the reference soil samples for field calibrations was far less than the amount of reference samples for the laboratory calibrations. The neighbouring locations of the reference samples with similar field conditions might have led to redundancies in the spectral field data leading to smaller field calibration errors. The greater data variability of the 97 samples measured in the laboratory, originating from 97 different sites in Northern Germany, could therefore not be modelled with equal satisfaction.

Next, the instrument calibrations were different for the NIR spectrometer used in this study. The laboratory instrument is calibrated yearly with a gold stamp, representing a 100% reflectance, and before each measurement series, with a background measurement, representing a zero reflectance measurement. The background spectrum is then subtracted from each sample spectrum stored. The field instrument is calibrated with four reference grey-scale standards, mentioned in section 3.2.2, before each measurement series. Each field

spectrum is transformed using an equation derived from the four reference standards. Additionally, the field instrument automatically collected dark spectra (zero reflectance) every 15 minutes. Overall, the field instrument was found to be much better calibrated than the laboratory instrument with respect to different grey scales (lab: two reference spectra for calibration; field: five reference spectra for calibration). As the soil colour originates mainly from soil organic matter, among other soil components, (Moritsuka et al., 2014) and SOC is the major component of soil organic matter, the better SOC calibration of the field instrument was assumed to be one of the explanatory factors for successful cross-validations with low calibration errors. However, the different resolutions and types of the NIR instruments as well as the reference analytic performed on SOC were found to have negligible effects on the SOC calibrations.

Generally, this work confirmed that laboratory- and field-based NIR measurements have great potential for assessing the organic carbon content in agricultural soils.

Even though promising results have been attained by the soil science community, the absence of standard procedures is a main reason why laboratory NIR measurements are not yet accepted as an equivalent alternative to conventional soil analytical techniques. As future work, the standardisation of the complete measurement procedure would improve the comparability of the spectral data collected from different NIR devices and also of spectral data collected over time from the same device. Moreover, research is needed to improve the performance of the VERIS shank and probe system, since on-line data acquisition regarding soil is a useful input for sensor-based precision agriculture and field experiments, for example, the application rate of herbicide could be adapted to heterogeneous SOC concentrations and for soil mapping needed for efficient soil monitoring in the frame of environmental monitoring. Moreover, the applicability of NIR for farming could be further tested by facilitating cost-effective devices with lower spectral resolutions and limited wavelength ranges.

7 References

Adamchuk, V.I., Rossel, R.A.V., Marx, D.B., Samal, A.K., 2011. Using targeted sampling to process multivariate soil sensing data. Geoderma 163, 63-73.

Adams, W.A., 1973. Effect of organic matter on bulk and true densities of some uncultivated podzolic soils. Journal of Soil Science 24, 10-17.

Anami, M.H., Marques, O.J., Goncalves, A.C.A., Folegatti, M.V., 2008. Nested sampling as a new approach to geostatistics applications compared with regular grid sampling. Central theme, technology for all: sharing the knowledge for development. Proceedings of the International Conference of Agricultural Engineering, XXXVII Brazilian Congress of Agricultural Engineering, International Livestock Environment Symposium - ILES VIII, Iguassu Falls City, Brazil, 31st August to 4th September, 2008, unpaginated.

Andrade-Sanchez, P., Upadhyaya, S.K., Jenkins, B.M., 2007. Development, construction and field evaluation of a soil compaction profile sensor. Transactions of the Asabe 50, 719-725.

Armstrong, P.R., Maghirang, E.B., Xie, F., Dowell, F.E., 2006. Comparison of dispersive and Fourier-transform NIR instruments for measuring grain and flour attributes. Applied Engineering in Agriculture 22, 453-457.

Barnes, R.J., Dhanoa, M.S., Lister, S.J., 1989. Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. Applied Spectroscopy 43, 772-777.

Barthes, B.G., Brunet, D., Ferrer, H., Chotte, J.L., Feller, C., 2006. Determination of total carbon and nitrogen content in a range of tropical soils using near infrared spectroscopy: influence of replication and sample grinding and drying. Journal of Near Infrared Spectroscopy 14, 341-348.

Bartholomeus, H.M., Schaepman, M.E., Kooistra, L., Stevens, A., Hoogmoed, W.B., Spaargaren, O.S.P., 2008. Spectral reflectance based indices for soil organic carbon quantification. Geoderma 145, 28-36.

Batjes, N.H., 1996. Total carbon and nitrogen in the soils of the world. European Journal of Soil Science 47, 151-163.

Bellon-Maurel, V., Fernandez-Ahumada, E., Palagos, B., Roger, J.-M., McBratney, A., 2010. Critical review of chemometric indicators commonly used for assessing the quality of the prediction of soil attributes by NIR spectroscopy. Trac-Trends in Analytical Chemistry 29, 1073-1081.

Bellon-Maurel, V., McBratney, A., 2011. Near-infrared (NIR) and mid-infrared (MIR) spectroscopic techniques for assessing the amount of carbon stock in soils - Critical review and research perspectives. Soil Biology & Biochemistry 43, 1398-1410.

Ben-Dor, E., Banin, A., 1995a. Near-infrared analysis (NIRA) as a method to simultaneously evaluate spectral featureless constituents in soils. Soil Science 159, 259-270.

Ben-Dor, E., Banin, A., 1995b. Quantitative analysis of convolved Thematic Mapper spectra of soils in the visible near-infrared and shortwave-infrared spectral regions. International Journal of Remote Sensing 16, 3509-3528.

Ben-Dor, E., Heller, D., Chudnovsky, A., 2008. A novel method of classifying soil profiles in the field using optical means. Soil Science Society of America Journal 72, 1113-1123.

Bengera, I., Norris, K.H., 1968. Determination of moisture content in soybeans by direct spectrophotometry. Israel Journal of Agricultural Research 18, 125-&.

Beste, A., 2002. Further development and improvement of spade diagnosis as field method for the evaluation of ecological significant structure parameters of soils under agricultural management. PhD-Thesis. Institute of Crop Management, University of Gießen.

Bird, S.B., Herrick, J.E., Wander, M.M., Wright, S.F., 2002. Spatial heterogeneity of aggregate stability and soil carbon in semi-arid rangeland. Environmental Pollution 116, 445-455.

Boucneau, G., Van Meirvenne, M., Thas, O., Hofman, G., 1998. Integrating properties of soil map delineations into ordinary kriging. European Journal of Soil Science 49, 213-229.

Brown, D.J., Shepherd, K.D., Walsh, M.G., Mays, M.D., Reinsch, T.G., 2006. Global soil characterization with VNIR diffuse reflectance spectroscopy. Geoderma 132, 273-290.

Brunet, D., Barthes, B.G., Chotte, J.-L., Feller, C., 2007. Determination of carbon and nitrogen contents in Alfisols, Oxisols and Ultisols from Africa and Brazil using NIRS analysis: Effects of sample grinding and set heterogeneity. Geoderma 139, 106-117.

Buening-Pfaue, H., 2003. Analysis of water in food by near infrared spectroscopy. Food Chemistry 82, 107-115.

Cahn, M.D., Hummel, J.W., Brouer, B.H., 1994. Spatial-analysis of soil fertility for site-specific crop management. Soil Science Society of America Journal 58, 1240-1248.

Calderon, F.J., Reeves, J.B., III, Foster, J.G., Clapham, W.M., Fedders, J.M., Vigil, M.F., Henry, W.B., 2007. Comparison of diffuse reflectance Fourier transform mid-infrared and near-infrared spectroscopy with grating-based near-infrared for the determination of fatty acids in forages. Journal of Agricultural and Food Chemistry 55, 8302-8309.

Callesen, I., Liski, J., Raulund-Rasmussen, K., Olsson, M.T., Tau-Strand, L., Vesterdal, L., Westman, C.J., 2003. Soil carbon stores in Nordic well-drained forest soils - relationships with climate and texture class. Global Change Biology 9, 358-370.

Cambardella, C.A., Moorman, T.B., Novak, J.M., Parkin, T.B., Karlen, D.L., Turco, R.F., Konopka, A.E., 1994. Field-scale variability of soil properties in Central Iowa soils. Soil Science Society of America Journal 58, 1501-1511.

Carmelino Hurtado, S.M., Silva, C.A., de Resende, A.V., Von Pinho, R.G., Borges Inacio, E.d.S., Higashikawa, F.S., 2009. Spatial variability of soil acidity attributes and the spatialization of liming requirement for corn. Ciencia E Agrotecnologia 33, 1351-1359.

Chang, C., Laird, D.A., Hurburgh, C.R., Jr., Chang, C.W., 2005. Influence of soil moisture on near-infrared reflectance spectroscopic measurement of soil properties. Soil Science 170, 244-255.

Chang, C.W., Laird, D.A., 2002. Near-infrared reflectance spectroscopic analysis of soil C and N. Soil Science 167, 110-116.

Chang, C.W., Laird, D.A., Mausbach, M.J., Hurburgh, C.R., 2001. Near-infrared reflectance spectroscopy-principal components regression analyses of soil properties. Soil Science Society of America Journal 65, 480-490.

Christy, C.D., 2008. Real-time measurement of soil attributes using on-the-go near infrared reflectance spectroscopy. Computers and Electronics in Agriculture 61, 10-19.

Christy, C.D., Drummond, P., Laird, D. A., 2003. An on-the-go spectral reflectance sensor for soil. Presentation at the 2003 ASAE Annual International Meeting Sponsored by ASAE Riviera Hotel and Convention Center, Las Vegas, Nevada, USA.

Ciais, P., Wattenbach, M., Vuichard, N., Smith, P., Piao, S.L., Don, A., Luyssaert, S., Janssens, I.A., Bondeau, A., Dechow, R., Leip, A., Smith, P.C., Beer, C., van der Werf, G.R., Gervois, S., Van Oost, K., Tomelleri, E., Freibauer, A., Schulze, E.D., Team, C.S., 2010. The European carbon balance. Part 2: croplands. Global Change Biology 16, 1409-1428.

Clark, R.N., 1999. Spectroscopy of rocks and minerals and principles of spectroscopy. In: Rencz, A.N. (Ed.), Manual of remote sensing. John Wiley & Sons, New York, pp. 3-58.

Conen, F., Zerva, A., Arrouays, D., Jolivet, C., Jarvis, P.G., Grace, J., Mencussini, M., 2005. The carbon balance of forest soils: detectability of changes in soil carbon stocks in temperate and Boreal forests. The Society for Experimental Biology, 235-249.

Couteau, O., Schaller, R., 2003. Mechanical spectroscopy of the interface in magnesium matrix compounds. Journal De Physique Iv 106, 63-68.

Cozzolino, D., Moron, A., 2003. The potential of near-infrared reflectance spectroscopy to analyse soil chemical and physical characteristics. Journal of Agricultural Science 140, 65-71.

Cozzolino, D., Moron, A., 2006. Potential of near-infrared reflectance spectroscopy and chemometrics to predict soil organic carbon fractions. Soil & Tillage Research 85, 78-85.

Craig, P.S., Shaw, T.J., Miller, P.L., Pellechia, P.J., Ferry, J.L., 2009. Use of Multiparametric Techniques To Quantify the Effects of Naturally Occurring Ligands on the Kinetics of Fe(II) Oxidation. Environmental Science & Technology 43, 337-342.

Dalal, R.C., Henry, R.J., 1986. Simultaneous determination of moisture, organic-carbon and total nitrogen by near-infrared reflectance spectrophotometry. Soil Science Society of America Journal 50, 120-123.

Davies, A.M.C., Fearn, T., 2006. Quantitative analysis via near infrared databases: comparison analysis using restructured near infrared and constituent data-deux (CARNAC-D). Journal of Near Infrared Spectroscopy 14, 403-411.

De Vos, B., Van Meirvenne, M., Quataert, P., Deckers, J., Muys, B., 2005. Predictive quality of pedotransfer functions for estimating bulk density of forest soils. Soil Science Society of America Journal 69, 500-510.

Deng, F., Minasny, B., Knadel, M., McBratney, A., Heckrath, G., Greve, M.H., 2013. Using Vis-NIR Spectroscopy for Monitoring Temporal Changes in Soil Organic Carbon. Soil Science 178, 389-399.

Donohue, S., Forristal, D., Donohue, L.A., 2013. Detection of soil compaction using seismic surface waves. Soil & Tillage Research 128, 54-60.

Duckworth, J., 2004. Mathematical data preprocessing. In: Roberts C.A., W.J., Jr., Reeves III, J.B. (Ed.), Near-Infrared Spectroscopy in Agriculture., Madison, ASA-CSSA-SSSA, pp. 115-132.

Dunn, B.W., Beecher, H.G., Batten, G.D., Ciavarella, S., 2002. The potential of near-infrared reflectance spectroscopy for soil analysis - a case study from the Riverine Plain of south-eastern Australia. Australian Journal of Experimental Agriculture 42, 607-614.

Efron, B., Tibshirani, R.J., 1994. An introduction to the Bootstrap. Chapmann & Hall/ CRC, Washington, D.C.

Esbensen, K.H., 2006. Multivariate analysis in practice: An introduction to multivariate data analysis and experimental design., Reprint of 2001 edition, fifth ed. CAMO Press, Norway.

Falloon, P., Smith, P., Bradley, R.I., Milne, R., Tomlinson, R., Viner, D., Livermore, M., Brown, T., 2006. RothCUK - a dynamic modelling system for estimating changes in soil C from mineral soils at 1-km resolution in the UK. Soil Use and Management 22, 274-288.

Fallou, F.A., 1862. Pedologie oder Allgemeine und Besondere Bodenkunde. Schönfeld Buchhandlung, Dresden.

Fystro, G., 2002. The prediction of C and N content and their potential mineralisation in heterogeneous soil samples using Vis-NIR spectroscopy and comparative methods. Plant and Soil 246, 139-149.

Gao, Y., Cui, L., Lei, B., Zhai, Y., Shi, T., Wang, J., Chen, Y., He, H., Wu, G., 2014. Estimating soil organic carbon content with visible-near-infrared (Vis-NIR) spectroscopy. Applied Spectroscopy 68, 712-722.

Garten, C.T., Wullschleger, S.D., 1999. Soil carbon inventories under a bioenergy crop (switchgrass): Measurement limitations. Journal of Environmental Quality 28, 1359-1365.

Gubler, A., 2011. Quantitative estimations of soil properties by Visible and Near Infrared Spectroscopy – applications for laboratory and field measurements. PhD-Thesis. Institute of Geography, University of Berne.

Gupta, R.K., Mostaghimi, S., McClellan, P.W., Alley, M.M., Brann, D.E., 1997. Spatial variability and sampling strategies for NO3-N, P, and K determinations for site-specific farming. Transactions of the Asae 40, 337-343.

Hampl, U., Kussel, N., 1994. Die Erweiterte Spatendiagnose. Unpublished.

Hbirkou, C., Paetzold, S., Mahlein, A.-K., Welp, G., 2012. Airborne hyperspectral imaging of spatial soil organic carbon heterogeneity at the field-scale. Geoderma 175, 21-28.

He, Y., Huang, M., Garcia, A., Hernandez, A., Song, H., 2007. Prediction of soil macronutrients content using near-infrared spectroscopy. Computers and Electronics in Agriculture 58, 144-153.

He, Y., Song, H., Pereira, A.G., Gomez, A.H., 2005. Measurement and analysis of soil nitrogen and organic matter content using near-infrared spectroscopy techniques. Journal of Zhejiang University (Science) 6B, 1081-1086.

Hedley, C.B., Payton, I.J., Lynn, I.H., Carrick, S.T., Webb, T.H., McNeill, S., 2012. Random sampling of stony and non-stony soils for testing a national soil carbon monitoring system. Soil Research 50, 18-29.

Hedley, M., Kusumo, B., Hedley, C., Tuohy, M., 2010. Field measurement of root density and soil organic carbon content using soil spectral reflectance. Proceedings of the 19th World Congress of Soil Science: Soil solutions for a changing world, Brisbane, Australia, 1-6 August 2010. Working Group 1.5 Soil sense: rapid soil measurements, 67-70.

Heise, H.M., Winzen, R., 2006. Chemometrics in near-infrared spectroscopy. In: Siesler H.W., O.Y., Kawata S., Heise H.M., (Ed.), Near-infrared spectroscopy: principles, instruments, applications. Germany: Wiley-VCH, pp. 125–162.

Hengl, T., Heuvelink, G.B.M., Stein, A., 2004. A generic framework for spatial prediction of soil variables based on regression-kriging. Geoderma 120, 75-93.

Heuscher, S.A., Brandt, C.C., Jardine, P.M., 2005. Using soil physical and chemical properties to estimate bulk density. Soil Science Society of America Journal 69, 51-56.

Hollis, J.M., Hannam, J., Bellamy, P.H., 2012. Empirically-derived pedotransfer functions for predicting bulk density in European soils. European Journal of Soil Science 63, 96-109.

Holmes, K.W., Wherrett, A., Keating, A., Murphy, D.V., 2011. Meeting bulk density sampling requirements efficiently to estimate soil carbon stocks. Soil Research 49, 680-695.

Huang, X., Senthilkurnar, S., Kravchenko, A., Thelen, K., Qi, J., 2007. Total carbon mapping in glacial till soils using near-infrared spectroscopy, Landsat imagery and topographical information. Geoderma 141, 34-42.

Hunt, G.R., Salisbury, J.W., 1970. Visible and near-infrared spectra of minerals and rocks. I. Silicate minerals. Modern Geology 1, 283–300.

IPCC, 2003. Good practice guidance for land use, land- use change and forestry, In: Penman, J., Gytarsky, M., Hiraishi, T., Krug, T., Kruger, D., Pipatti, R., Buendia, L., Miwa, K., Ngara, T., Tanabe, K., Wagner, F. (Eds.), Hayama, Japan.

Javed, I., Thomasson, J.A., Jenkins, J.N., Owens, P.R., Whisler, F.D., 2005. Spatial variability analysis of soil physical properties of alluvial soils. Soil Science Society of America Journal 69, 1338–1350.

Jobbagy, E.G., Jackson, R.B., 2000. The vertical distribution of soil organic carbon and its relation to climate and vegetation. Ecological Applications 10, 423-436.

Jordanova, N., Jordanova, D., Tsacheva, T., 2008. Application of magnetometry for delineation of anthropogenic pollution in areas covered by various soil types. Geoderma 144, 557-571.

Kawano, S., Abe, H., 1995. Development of a calibration equation with temperature compensation for determing the Brix value in intact peaches. Journal of Near Infrared Spectroscopy 3, 211-218.

Kerry, R., Oliver, M.A., 2004. Average variograms to guide soil sampling. International Journal of Applied Earth Observation and Geoinformation 5, 307-325.

Kitchen, N.R., Sudduth, K.A., Drummond, S.T., Scharf, P.C., Palm, H., Shannon, D.K., 2008. Assessing intra- and inter-field variability of corn nitrogen fertilizer need using ground-based reflectance sensors. Proceedings of the 9th International Conference on Precision Agriculture, Denver, Colorado, USA, 20-23 July, 2008. Abstract.

Knadel, M., Thomsen, A., Greve, M.H., 2011. Multisensor on-the-go mapping of soil organic carbon content. Soil Science Society of America Journal 75, 1799-1806.

Krebs, C.J., 1999. Ecological methodology. Addison Wesley Longman, Menlo Park, CA.

Kretzschmar, R., 1986. Kulturtechnisch-Bodenkundliches Praktikum. Institut für Wasserwirtschaft und Landschaftökologie, Christian-Albrechts-Universität Kiel.

Krishnan, P., Alexander, J.D., Butler, B.J., Hummel, J.W., 1980. Reflectance technique for predicting soil organic matter. Soil Science Society of America Journal 44, 1282-1285.

Kuang, B., Mahmood, H.S., Quraishi, M.Z., Hoogmoed, W.B., Mouazen, A.M., van Hentent, E.J., 2012. Sensing soil properties in the laboratory, in situ and on-line: A review. In: Sparks, D.L. (Ed.), Advances in Agronomy, Vol 114, pp. 155-223.

Kuang, B., Mouazen, A.M., 2011. Calibration of visible and near infrared spectroscopy for soil analysis at the field scale on three European farms. European Journal of Soil Science 62, 629-636.

Kuang, B., Mouazen, A.M., 2013. Non-biased prediction of soil organic carbon and total nitrogen with vis-NIR spectroscopy, as affected by soil moisture content and texture. Biosystems Engineering 114, 249-258.

Kumar, S., 2015. Estimating spatial distribution of soil organic carbon for the Midwestern United States using historical database. Chemosphere 127, 49-57.

Kumar, S., La, R., 2011. Mapping the organic carbon stocks of surface soils using local spatial interpolator. Journal of Environmental Monitoring 13, 3128-3135.

Lal, R., 2001. World cropland soils as a source or sink for atmospheric carbon, In: Sparks, D.L. (Ed.), Advances in Agronomy, Vol 71, pp. 145-191.

Lal, R., 2004. Agricultural activities and the global carbon cycle. Nutrient Cycling in Agroecosystems 70, 103-116.

Lal, R., 2008. Carbon sequestration. Philosophical Transactions of the Royal Society B-Biological Sciences 363, 815-830.

Leifeld, J., Bassin, S., Fuhrer, J., 2005. Carbon stocks in Swiss agricultural soils predicted by land-use, soil characteristics, and altitude. Agriculture Ecosystems & Environment 105, 255-266.

Li, Q.-q., Yue, T.-x., Wang, C.-q., Zhang, W.-j., Yu, Y., Li, B., Yang, J., Bai, G.-c., 2013. Spatially distributed modeling of soil organic matter across China: An application of artificial neural network approach. Catena 104, 210-218.

Lischer, P., 1993. Source, propagation and control of errors in soil monitoring. In: Schulin, R., Desaules, A., Webster, R. (Ed.), Soil monitoring: Early detection and surveying of soil contamination and degradation, pp. 277-294.

Liu, Z.-P., Shao, M.-A., Wang, Y.-Q., 2012. Estimating soil organic carbon across a large-scale region: A state-space modeling approach. Soil Science 177, 607-618.

Lorenz, K., Lal, R., 2005. The depth distribution of soil organic carbon in relation to land use and management and the potential of carbon sequestration in subsoil horizons, In: Sparks, D.L. (Ed.), Advances in Agronomy, Vol 88, pp. 35-66.

Ludwig, B., Khanna, P.K., Bauhus, J., Hopmans, P., 2002. Near infrared spectroscopy of forest soils to determine chemical and biological properties related to soil sustainability. Forest Ecology and Management 171, 121-132.

Madari, B.E., Reeves, J.B., Coelho, M.R., Machado, P., De-Polli, H., 2005. Mid- and near-infrared spectroscopic determination of carbon in a diverse set of soils from the Brazilian National Soil Collection. Spectroscopy Letters 38, 721-740.

Malhi, S.S., Grant, C.A., Johnston, A.M., Gill, K.S., 2001. Nitrogen fertilization management for no-till cereal production in the Canadian Great Plains: a review. Soil & Tillage Research 60, 101-122.

Malley, D.F., Yesmin, L., Eilers, R.G., 2002. Rapid analysis of hog manure and manure-amended soils using near-infrared spectroscopy. Soil Science Society of America Journal 66, 1677-1686.

Marin-Gonzalez, O., Kuang, B., Quraishi, M.Z., Angel Munoz-Garcia, M., Mouazen, A.M., 2013. On-line measurement of soil properties without direct spectral response in near infrared spectral range. Soil & Tillage Research 132, 21-29.

Martens, H., Naes, T., 1989. Multivariate calibration. Wiley, New York.

Martin, P.D., Malley, D.F., Manning, G., Fuller, L., 2002. Determination of soil organic carbon and nitrogen at the field level using near-infrared spectroscopy. Canadian Journal of Soil Science 82, 413-422.

McArthur, L., Greensill, C., 2007. Comparison of two NIR systems for quantifying kaolinite inWeipa bauxites. Measurement Science & Technology 18, 3463-3470.

McBratney, A.B., Minasny, B., Rossel, R.V., 2006. Spectral soil analysis and inference systems: A powerful combination for solving the soil data crisis. Geoderma 136, 272-278.

McCarty, G.W., Reeves, J.B., Reeves, V.B., Follett, R.F., Kimble, J.M., 2002. Mid-infrared and near-infrared diffuse reflectance spectroscopy for soil carbon measurement. Soil Science Society of America Journal 66, 640-646.

Meersmans, J., van Wesemael, B., Goidts, E., van Molle, M., De Baets, S., De Ridder, F., 2011. Spatial analysis of soil organic carbon evolution in Belgian croplands and grasslands, 1960-2006. Global Change Biology 17, 466-479.

Michel, K., Terhoeven-Urselmans, T., Nitschke, R., Steffan, P., Ludwig, B., 2009. Use of near- and mid-infrared spectroscopy to distinguish carbon and nitrogen originating from char and forest-floor material in soils. Journal of Plant Nutrition and Soil Science-Zeitschrift Fur Pflanzenernaehrung und Bodenkunde 172, 63-70.

Miltz, J., Don, A., 2012. Optimising sample preparation and near infrared spectra measurements of soil samples to calibrate organic carbon and total nitrogen content. Journal of Near Infrared Spectroscopy 20, 695-706.

Mishra, U., Torn, M.S., Masanet, E., Ogle, S.M., 2012. Improving regional soil carbon inventories: Combining the IPCC carbon inventory method with regression kriging. Geoderma 189, 288-295.

Mondini, C., Coleman, K., Whitmore, A.P., 2012. Spatially explicit modelling of changes in soil organic C in agricultural soils in Italy, 2001-2100: Potential for compost amendment. Agriculture Ecosystems & Environment 153, 24-32.

Moritsuka, N., Matsuoka, K., Katsura, K., Sano, S., Yanai, J., 2014. Soil color analysis for statistically estimating total carbon, total nitrogen and active iron contents in Japanese agricultural soils. Soil Science and Plant Nutrition 60, 475-485.

Moron, A., Cozzolino, D., 2002. Application of near infrared reflectance spectroscopy for the analysis of organic C, total N and pH in soils of Uruguay. Journal of near Infrared Spectroscopy 10, 215-221.

Moron, A., Cozzolino, D., 2004. Determination of potentially mineralizable nitrogen and nitrogen in particulate organic matter fractions in soil by visible and near-infrared reflectance spectroscopy. Journal of Agricultural Science 142, 335-343.

Mouazen, A.M., 2009. The future for on-line measurement of soil properties with sensor fusion systems. Landwards 64, 14-16.

Mouazen, A.M., De Baerdemaeker, J., Ramon, H., 2006. Effect of wavelength range on the measurement accuracy of some selected soil constituents using visual-near infrared spectroscopy. Journal of Near Infrared Spectroscopy 14, 189-199.

Mouazen, A.M., Karoui, R., Baerdemaeker, J.d., Ramon, H., de Baerdemaeker, J., 2007a. Classification of soil texture classes for on-the-go management of soil VIS-NIR spectra. Papers presented at the 6th European Conference on Precision Agriculture, Skiathos, Geece, 3-6 June 2007.

Mouazen, A.M., Kuang, B., De Baerdemaeker, J., Ramon, H., 2010. Comparison among principal component, partial least squares and back propagation neural network analyses for accuracy of measurement of selected soil properties with visible and near infrared spectroscopy. Geoderma 158, 23-31.

Mouazen, A.M., Maleki, M.R., De Baerdemaeker, J., Ramon, H., 2007b. On-line measurement of some selected soil properties using a VIS-NIR sensor. Soil & Tillage Research 93, 13-27.

Munoz, J.D., Kravchenko, A., 2011. Soil carbon mapping using on-the-go near infrared spectroscopy, topography and aerial photographs. Geoderma 166, 102-110.

Naes, T., Isaksson, T., 1989. Selection of samples for calibration in near-infrared spectroscopy .1. General-principles illustrated by example. Applied Spectroscopy 43, 328-335.

Nash, E., Wiebensohn, J., Nikkila, R., Vatsanidou, A., Fountas, S., Bill, R., 2011. Towards automated compliance checking based on a formal representation of agricultural production standards. Computers and Electronics in Agriculture 78, 28-37.

Nduwamungu, C., Ziadi, N., Parent, L.-E., Tremblay, G.F., Thuries, L., 2009a. Opportunities for, and limitations of, near infrared reflectance spectroscopy applications in soil analysis: A review. Canadian Journal of Soil Science 89, 531-541.

Nduwamungu, C., Ziadi, N., Tremblay, G.F., Parent, L.-E., 2009b. Near-infrared reflectance spectroscopy prediction of soil properties: Effects of sample cups and preparation. Soil Science Society of America Journal 73, 1896-1903.

Nocita, M., Kooistra, L., Bachmann, M., Mueller, A., Powell, M., Weel, S., 2011. Predictions of soil surface and topsoil organic carbon content through the use of laboratory and field spectroscopy in the Albany Thicket Biome of Eastern Cape Province of South Africa. Geoderma 167-68, 295-302.

Nyssen, J., Temesgen, H., Lemenih, M., Zenebe, A., Haregeweyn, N., Haile, M., 2008. Spatial and temporal variation of soil organic carbon stocks in a lake retreat area of the Ethiopian Rift Valley. Geoderma 146, 261-268.

Ogle, S.M., Breidt, F.J., Paustian, K., 2006. Bias and variance in model results associated with spatial scaling of measurements for parameterization in regional assessments. Global Change Biology 12, 516-523.

Ogunwole, J.O., Obidike, E.O., Timm, L.C., Odunze, A.C., Gabriels, D.M., 2014. Assessment of spatial distribution of selected soil properties using geospatial statistical tools. Communications in Soil Science and Plant Analysis 45, 2182-2200.

Oliver, M.A., Webster, R., 2014. A tutorial guide to geostatistics: Computing and modelling variograms and kriging. Catena 113, 56-69.

Oltra-Carrio, R., Baup, F., Fabre, S., Fieuzal, R., Briottet, X., 2015. Improvement of soil moisture retrieval from hyperspectral VNIR-SWIR data using clay content information: From laboratory to field experiments. Remote Sensing 7, 3184-3205.

Orth, A., 1877. Die Schwarzerde und ihre Bedeutung für die Kultur (The black soil and its impact on culture). Die Natur 26, 36–38 (in German).

Otto, F., 1997. L1-contraction and uniqueness for unstationary saturated-unsaturated water flow in porous media. Adv. Math. Sci. Appl. 7, 537–553.

Peck, T.R., Soltanpour, P.N., 1990. The principles of soil testing. In: Westermann, R. L. (Ed.), Soil testing and plant analysis, 3rd Ed., pp. 1-9.

Peng, X., Shi, T., Song, A., Chen, Y., Gao, W., 2014. Estimating Soil Organic Carbon Using VIS/NIR Spectroscopy with SVMR and SPA Methods. Remote Sensing 6, 2699-2717.

Pettinelli, E., Vannaroni, G., Di Pasquo, B., Mattei, E., Di Matteo, A., De Santis, A., Annan, P.A., 2007. Correlation between near-surface electromagnetic soil parameters and early-time GPR signals: An experimental study. Geophysics 72, A25-A28.

Pirie, A., Singh, B., Islam, K., 2005. Ultra-violet, visible, near-infrared, and mid-infrared diffuse reflectance spectroscopic techniques to predict several soil properties. Australian Journal of Soil Research 43, 713-721.

Quraishi, M.Z., Mouazen, A.M., 2013. Calibration of an on-line sensor for measurement of topsoil bulk density in all soil textures. Soil & Tillage Research 126, 219-228.

Raper, R.L., Erbach, D.C., 1987. Bulk-density measurement variability with core samplers. Transactions of the Asae 30, 878-881.

Rawls, W.J., 1983. Estimating soil bulk-density from particle-size analysis and organic matter content. Soil Science 135, 123-125.

Reeves, J., McCarty, G., Mimmo, T., 2002. The potential of diffuse reflectance spectroscopy for the determination of carbon inventories in soils. Environmental Pollution 116, S277-S284.

Reeves, J.B., Zapf, C.M., 1999. Spectral library searching: Mid-infrared versus near-infrared spectra for classification of powdered food ingredient. Applied Spectroscopy 53, 836-844.

Rinnan, R., Rinnan, A., 2007. Application of near infrared reflectance (NIR) and fluorescence spectroscopy to analysis of microbiological and chemical properties of arctic soil. Soil Biology & Biochemistry 39, 1664-1673.

Rodionov, A., Welp, G., Damerow, L., Berg, T., Amelung, W., Paetzold, S., 2015. Towards on-the-go field assessment of soil organic carbon using Vis-NIR diffuse reflectance spectroscopy: Developing and testing a novel tractor-driven measuring chamber. Soil & Tillage Research 145, 93-102.

Rossel, R.A.V., Fouad, Y., Walter, C., 2008. Using a digital camera to measure soil organic carbon and iron contents. Biosystems Engineering 100, 149-159.

Rueth, B., Lennartz, B., 2008. Spatial variability of soil properties and rice yield along two catenas in southeast China. Pedosphere 18, 409-420.

Russell, C.A., 2003. Sample preparation and prediction of soil organic matter properties by near infra-red reflectance spectroscopy. Communications in Soil Science and Plant Analysis 34, 1557-1572.

Santra, P., Chopra, U.K., Chakraborty, D., 2008. Spatial variability of soil properties and its application in predicting surface map of hydraulic parameters in an agricultural farm. Current Science 95, 937-945.

Sawyer, J.E., 1994. Concepts of variable-rate technology with considerations for fertilizer application. Journal of Production Agriculture 7, 195-201.

Schoening, I., Totsche, K.U., Koegel-Knabner, I., 2006. Small scale spatial variability of organic carbon stocks in litter and solum of a forested Luvisol. Geoderma 136, 631-642.

Schrumpf, M., Schulze, E.D., Kaiser, K., Schumacher, J., 2011. How accurately can soil organic carbon stocks and stock changes be quantified by soil inventories? Biogeosciences 8, 1193-1212.

Senft, F., 1857. Lehrbuch der forstlichen Geognosie, Bodenkunde und Chemie. Manke, Jena.

Seong, K.-C., Iyo, C., Kawano, S., 1999. Development of NIR calibration equation with temperature compensation for determining the brix value in intact strawberrys. Journal of the Korean Society for Horticultural Science 40, 673-677.

Sequeira, C.H., Wills, S.A., Grunwald, S., Ferguson, R.R., Benham, E.C., West, L.T., 2014. Development and Update Process of VNIR-Based Models Built to Predict Soil Organic Carbon. Soil Science Society of America Journal 78, 903-913.

Sethuramasamyraja, B., Adamchuk, V.I., Dobermann, A., Marx, D.B., Jones, D.D., Meyer, G.E., 2008. Agitated soil measurement method for integrated on-the-go mapping of soil pH, potassium and nitrate contents. Computers and Electronics in Agriculture 60, 212-225.

Shenk, J.S., Westerhaus, M.O., 1991. New standardization and calibration procedures for NIRS analytical systems. Crop Science 31, 1694-1696.

Shepherd, K.D., Walsh, M.G., 2002. Development of reflectance spectral libraries for characterization of soil properties. Soil Science Society of America Journal 66, 988-998.

Shepherd, K.D., Walsh, M.G., 2007. Infrared spectroscopy - enabling an evidence-based diagnostic surveillance approach to agricultural and environmental management in developing countries. Journal of near Infrared Spectroscopy 15, 1-19.

Shibusawa, S., Anom, W.S.I.M., Sato, H., Sasao, A., Hirako, S., Otomo, A., Blackmore, S., 2000a. On-line real-time soil spectrophotometer. Proceedings of the 5th International Conference on Precision Agriculture, Bloomington, Minnesota, USA. 16-19 July, 2000, pp. 1-13.

Shibusawa, S., Sato, H., Anom, S.W.I.M., Sasao, A., Sakai, K., Hirako, S., 2000b. Stability of soil reflectance measurement by the real-time spectrophotometer. Proceedings of the 5th International Conference on Precision Agriculture, Bloomington, Minnesota, USA. 16-19 July, 2000, pp. 1-9.

Sleutel, S., De Neve, S., Beheydt, D., Li, C., Hofman, G., 2006. Regional simulation of long-term organic carbon stock changes in cropland soils using the DNDC model: 2. Scenario analysis of management options. Soil Use and Management 22, 352-361.

Smith, C.A.S., Lobb, D.A., Monreal, C.M., 2005. Estimating regional soil organic carbon stocks. Canadian Journal of Soil Science 85, 463-465.

Smith, P., 2004. How long before a change in soil organic carbon can be detected? Global Change Biology 10, 1878-1883.

Sorensen, L.K., Dalsgaard, S., 2005. Determination of clay and other soil properties by near infrared spectroscopy. Soil Science Society of America Journal 69, 159-167.

Srinivasan, M.S., Bryant, R.B., Callahan, M.P., Weld, J.L., 2006. Manure management and nutrient loss under winter conditions: A literature review. Journal of Soil and Water Conservation 61, 200-209.

Stenberg, B., 2010. Effects of soil sample pretreatments and standardised rewetting as interacted with sand classes on Vis-NIR predictions of clay and soil organic carbon. Geoderma 158, 15-22.

Stevens, A., van Wesemael, B., Bartholomeus, H., Rosillon, D., Tychon, B., Ben-Dor, E., 2008. Laboratory, field and airborne spectroscopy for monitoring organic carbon content in agricultural soils. Geoderma 144, 395-404.

Sudduth, K.A., Drummond, S.T., Kitchen, N.R., 2001. Accuracy issues in electromagnetic induction sensing of soil electrical conductivity for precision agriculture. Computers and Electronics in Agriculture 31, 239-264.

Sudduth, K.A., Hummel, J.W., 1993.Portable, near-infrared spectophotometer for rapid soil analysis. Transactions of the Asae 36, 185-193.

Szalai, Z., Kiss, K., Jakab, G., Sipos, P., Belucz, B., Nemeth, T., 2013. The use of UV-VIS-NIR reflectance spectroscopy to identify iron minerals. Astronomische Nachrichten 334, 940-943.

Tekin, Y., Tumsavas, Z., Mouazen, A.M., 2010. Effect of moisture content on prediction of organic carbon and pH using visible and near infrared spectroscopy.

Tekin, Y., Tumsavas, Z., Mouazen, A.M., 2014. Comparing the artificial neural network with parcial least squares for prediction of soil organic carbon and PH at different moisture content levels using visible and near-infrared spectroscopy. Revista Brasileira De Ciencia Do Solo 38, 1794-1804.

Terhoeven-Urselmans, T., Schmidt, H., Joergensen, R.G., Ludwig, B., 2008. Usefulness of near-infrared spectroscopy to determine biological and chemical soil properties: Importance of sample pre-treatment. Soil Biology & Biochemistry 40, 1178-1188.

Tornquist, C.G., Mielniczuk, J., Pellegrino Cerri, C.E., 2009. Modeling soil organic carbon dynamics in Oxisols of Ibiruba (Brazil) with the Century Model. Soil & Tillage Research 105, 33-43.

Troeh, F.R., Hobbs, J.A., Donahue, R.L., 2004. Soil and water conservation for productivity and environmental protection. 4th Ed. Prentice Hall, Upper Saddle River, NJ.

Udelhoven, T., Emmerling, C., Jarmer, T., 2003. Quantitative analysis of soil chemical properties with diffuse reflectance spectrometry and partial least-square regression: A feasibility study. Plant and Soil 251, 319-329.

VandenBygaart, A.J., 2006. Monitoring soil organic carbon stock changes in agricultural landscapes: Issues and a proposed approach. Canadian Journal of Soil Science 86, 451-463.

VandenBygaart, A.J., Gregorich, E.G., Angers, D.A., McConkey, B.G., 2007. Assessment of the lateral and vertical variability of soil organic carbon. Canadian Journal of Soil Science 87, 433-444.

Vandervaere, J.P., Vauclin, M., Haverkamp, R., Cuenca, R.H., 1994. Error analysis in estimating soil-water balance of irrigated fields during the Efeda experiment. 1. Local standpoint. Journal of Hydrology 156, 351-370.

Viscarra Rossel, R.A., Behrens, T., 2010. Using data mining to model and interpret soil diffuse reflectance spectra. Geoderma 158, 46-54.

Viscarra Rossel, R.A., Cattle, S.R., Ortega, A., Fouad, Y., 2009. In situ measurements of soil colour, mineral composition and clay content by vis-NIR spectroscopy. Geoderma 150, 253-266.

Viscarra Rossel, R.A., Fouad, Y., Walter, C., 2008. Using a digital camera to measure soil organic carbon and iron contents. Biosystems Engineering 100, 149-159.

Viscarra Rossel, R.A., Walvoort, D.J.J., McBratney, A.B., Janik, L.J., Skjemstad, J.O., 2006. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. Geoderma 131, 59-75.

Wang, J., Wu, Y., Fu, Q., 1999. Study on spatial variability in conductivity of the coastal saline soils. Journal of Zhejiang Agricultural University 25, 139-142.

Wang, Y., Huang, T., Liu, J., Lin, Z., Li, S., Wang, R., Ge, Y., 2015. Soil pH value, organic matter and macronutrients contents prediction using optical diffuse reflectance spectroscopy. Computers and Electronics in Agriculture 111, 69-77.

Webster, R., Oliver, M.A., 1992. Sample adequately to estimate variograms of soil properties. Journal of Soil Science 43, 177-192.

Wessolek, G., Kaupenjohann, M., Dominik, P., Ilg, K., Schmitt, A., Zeitz, J., Gahre, F., Schulz, E., Ellerbrock, R.H., Utermann, J., Düwel, O., Siebner, C., 2008. Ermittlung von Optimalgehalten an organischer Substanz landwirtschaftlich genutzter Böden nach § 17 (2) Nr. 7 BBodSchG. Umweltbundesamt, Berlin.

Wetterlind, J., Stenberg, B., 2010. Near-infrared spectroscopy for within-field soil characterization: small local calibrations compared with national libraries spiked with local samples. European Journal of Soil Science 61, 823-843.

Wight, J.P., Ashworth, A.J., Allen, F.L., 2016. Organic substrate, clay type, texture, and water influence on NIR carbon measurements. Geoderma 261, 36-43.

Wold, S., Ruhe, A., Wold, H., Dunn, W.J., 1984. The collinearity problem in linear regression - the partial least squares (PLS) approach to generalized inverses. Siam Journal on Scientific and Statistical Computing 5, 735-743.

Wold, S., Trygg, J., Berglund, A., Antti, H., 2001. Some recent developments in PLS modeling. Chemometrics and Intelligent Laboratory Systems 58, 131-150.

Worsham, L., Markewitz, D., Nibbelink, N., 2010. Incorporating spatial dependence into estimates of soil carbon contents under different land covers. Soil Science Society of America Journal 74, 635-646.

Wu, C.-Y., Jacobson, A.R., Laba, M., Baveye, P.C., 2009. Alleviating Moisture Content Effects on the Visible Near-Infrared Diffuse-Reflectance Sensing of Soils. Soil Science 174, 456-465.

Youden, W.J., Mehlich, A., 1938. Selection of efficient methods for soil sampling. Proceedings. Soil Science Society of America 2, 399-399.

Zornoza, R., Guerrero, C., Mataix-Solera, J., Scow, K.M., Arcenegui, V., Mataix-Beneyto, J., 2008. Near infrared spectroscopy for determination of various physical, chemical and biochemical properties in Mediterranean soils. Soil Biology & Biochemistry 40, 1923-1930.

8 Curriculum Vitae

Personal data

Name	Jasmin Aymone Miltz
Date and place of birth	18th November 1975 in Celle (Germany)
Nationality	German

Education

2010 – 2016	Doctoral candidate, Institute of Geoecology, Technical University of Braunschweig
2004 – 2005	Qualification in International Management, LEB Hannover and Forschungszentrum Jülich GmbH, Berlin
1995 – 2001	Studies in geology and palaeontology, Leibniz University of Hannover
1995	Abitur, Hermann-Billung-Gymnasium in Celle

Work experience

2013 – 2016	Visiting scientist, Thünen Institute, Braunschweig
2010 – 2012	Research assistant, Thünen Institute, Braunschweig,
2007 – 2009	Soil scientist, leader of science workshops, phaeno gGmbH, Wolfsburg
2006 – 2007	Freelance consulting geologist, Geopark Harz, Braunschweiger Land, Ostfalen
2005 – 2006	Freelance geologist, Landesmuseum Hannover
2001 – 2003	Research assistant, Institute of Soil Science, University of Hannover

Publications

The laboratory aspect of NIRS, as one subject of this thesis, has been published in the Journal of Near Infrared Spectroscopy as follows:

J. Miltz and A. Don (2012):
„Optimising sample preparation and near infrared spectra measurements of soil samples to calibrate organic carbon and total nitrogen content". Journal of Near Infrared Spectroscopy, Volume 20, Issue 6, Pages: 695-706.

9 Acknowledgements

A thousand thanks to the many persons who contributed to the present work.

I thank Professor Rolf Nieder and Professor Heinz Flessa for helpful discussions and advice.

I am grateful to Dr Axel Don for his generous support and supervision in the NIRVANA project over the last years and for the many instances of amusing field moments summed up with ‘Wenn es läuft, läufts!’

I thank Eric Lund and Chase Maxton (VERIS Technologies, Inc., Kansas, USA) for introducing me to their shank and probe system and for answering numerous emails regarding ‘how to run the measurements’.

My gratitude goes to Marc Jantz and Dirk Lempio for help in repairing the field devices and Andreas Laggner for help in solving problems concerning the field software.

Thanks to Dr Andreas Hoffmann (LUFA Nord-West) for providing the soil samples, as well as Miriam Kuhlmann, Ute Riess and Silke Weis for spectra acquisition in the laboratory.

I sincerely thank Dr Klemens Irmer (Bruker Optik GmbH) for providing different spectrometer ring cups for the comparisons and Dr Heinrich Prüfer as well as Annika Studt (SensoLogic GmbH) for helpful discussions about calibration models.

Thanks also goes to Dr Jens Schumacher (Friedrich-Schiller-University, Jena, Germany) and Dr Roland Fuss (Thünen Institute) for help in programming the R-scripts.

I gratefully acknowledge all friends and colleagues from the office ‘Dokraum33’ (Caroline Buchen, Greta Roth, Dr Ulrike Wolf and Nicolas Ruoss) and from the Thünen Institute (Dr Michaela Busch, Peter Braunisch, Dr Stefan Burkart, Dr Marianna Deppe, Dr Anja-Bettina Dohrmann, Jennifer Ehe, Dr Nina Eibisch, Sven Evertsbusch, Anke Führer, Dr Andreas Gensior, Dr Anette Giesemann, Dr Mirjam Helfrich, Michael Hemkemeyer, Dr Sebastian Klimek, Dr Werner Kutsch, Daniela Müller, Dr Katharina Plassmann, Dr Christopher Poeplau, Ute Riess, Dr Lena Rohe, Ute Tambor and Claudia Wiese) for assistance in the field and/or laboratory and/or for fruitful discussions or for simply good times.

I would like to express my immense gratitude to my parents for their love and trust in me throughout my life; you strengthen me and bring and share happiness.

And last but certainly not least, I thank my husband and my son for their love, patience, mental support, understanding and many funny moments.